

ESCUELA POLITÉCNICA SUPERIOR



Análisis de sentimientos acerca de marcas mediante Tweets

Autor: Laura Gallego-Preciado Cobos

Tutor: Telmo Zarraonandia

Leganés, noviembre 2013





Título: Análisis de sentimientos acerca de marcas mediante Tweets

Autor: Laura Gallego-Preciado Cobos

Director: Telmo Zarraonandia

EL TRIBUNAL

Presidente: _____

Vocal: _____

Secretario: _____

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 19 de noviembre del 2013 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE





Agradecimientos

En primer lugar, tengo que dar gracias a mis padres, Bernabé y Maricarmen, agradecerles que siempre han estado ahí cuando les necesitaba y cuando no, también agradecerles el esfuerzo que les ha supuesto y agradecerles la confianza que depositaron en mí sabiendo que iba a ser capaz de terminar mi carrera cuando otros no apostaban por ello. Por esto y muchas otras cosas quiero agradecerleselo.

En segundo lugar a mi tía, Teresa, que siempre ha estado ahí apoyándome y ayudándome de la mejor forma que ella sabe, queriéndome y haciéndome la comida.

Seguido tengo que mencionar a Carlos, junto con mis padres y mi abuela ha sido el que más me ha tenido que soportar (en cierto modo es lo que le toca).

Por otro lado tengo que hacer una mención general a toda mi familia, mis tíos: Fidel, Dorita, Pepe, Patricia, Amando, Nati, Manolo y Gema que son los que me han dado consejos y ayudado en la parte más “teórica” de la carrera y por otro lado a mis primos: Miguel, Emi, Pilar, Carol, Jesús, Mari José, Eduardo, Diego, Maneni, Gema, Victoria y Daniel que son los que me han ayudado en la parte más “práctica” de la carrera. Por tanto muchas gracias a toda mi familia que ha estado ahí siempre para ayudarme.

Gracias a mi amiga Laura que me ha tenido que aguantar mucho durante estos 6 años que llevamos juntas, algunos más de los que llevo en esta carrera.

Cómo olvidarme de mis compis, aunque aparezcáis aquí abajo no sois menos importantes, a los que quiero mucho y nunca me olvidare de ellos. Coral, Roberto, Fran, Adrián, Alberto, Sofía o José entre muchos otros, los cuales me han apoyado y brindado su amistad y que sin su ayuda hubiera sido casi imposible haberme sacado esta carrera.

Por otro lado a mis seres queridos que ya no están aquí, que de una forma u otra también me han ayudado. Mis abuelos Fidel y Pilar, mi abuelo Manuel, mi primo David, mi tía Amanda.

Para terminar, me gustaría agradecer a Telmo que quisiera ser mi director del proyecto y espero que no le haya creado muchos quebraderos de cabeza y que el resultado final de mi proyecto fin de carrera sea de su agrado.

Para terminar y que quede más claro:

01001101 01110101 01100011 01101000 01100001 01110011 00100000
01100111 01110010 01100001 01100011 01101001 01100001 01110011 00100000
01100001 00100000 01110100 01101111 01100100 01101111 01110011 00101110

(“**Muchas gracias a todos**”)

Laura Gallego-Preciado Cobos





Resumen

Las redes sociales son un fenómeno en auge. En los países occidentales se calcula que más 30% de la población se comunica habitualmente utilizando esta tecnología. En el caso de España las redes sociales más populares son *Facebook*, *Tuenti* y *Twitter*. Esta última red, *Twitter*, ofrece una funcionalidad algo distinta al resto, puesto que el usuario únicamente dispone de 140 caracteres para escribir sus mensajes, en los que habitualmente se tratan y discuten temas de actualidad. Las opiniones vertidas en los *tweets* pueden ser de gran utilidad para las empresas y marcas, ya que pueden servir para obtener una valoración que los usuarios tienen de los servicios ofrecidos por ellas. Siguiendo esta idea en este proyecto de fin de carrera se plantea el desarrollo de una aplicación de escritorio que tendrá como objetivo ayudar a un **Community Manager** a **conocer** mejor qué **opinan** los usuarios de **Twitter** sobre la empresa para la que él trabaja, **"su marca"**.

Palabras clave: **Community Manager**, **opinión**, **marca**, **Twitter**.





Abstract

Social networks are a growing phenomenon. More than the 30% of the population that lives in western countries use these kinds of networks to communicate with family and friends every day. In Spain, the most social networks used by the population are Facebook, Tuenti and Twitter. Despite being known as social networks, each one is focused on a different sector of the population, because in each one you can find various functions. But in this case, the social network we are going to focus on is **Twitter**, this social network is different from the rest, since you only have 140 characters to write what you want, but without going over. In the tweets people criticize, think, discuss, and talk about current issues. This social network has been widely accepted, because you can express yourself with those concrete numbers of characters. Throughout this paper, I will explain the process, which has continued to develop the desktop application. The desktop application will aim to assist the **Community Manager** to **know** what users in **Twitter say** about the company for he works "**his brand**".

Keywords: **Community Manager, opinion, brand, Twitter.**





Índice de contenidos

Contents

1	INTRODUCCIÓN.....	20
1.1	Contexto.....	20
1.2	Facebook	21
1.3	tuenti	22
1.4	twitter.....	23
1.4.1	Dispositivos donde puede utilizarse	24
1.5	El reto.....	24
1.6	Soluciones propuestas	26
1.7	Objetivos	27
1.7.1	Consideraciones previas.....	27
1.7.2	Objetivos	28
1.8	Estructura del documento	28
2	ESTADO DEL ARTE.....	33
2.1	Análisis de sentimientos.....	33
2.1.1	Definición	33
2.1.2	¿Para qué se utiliza?	38
2.2	Twitter	38
2.2.1	¿Qué es Twitter?.....	38
2.2.2	Historia de Twitter	38
2.2.3	Conceptos relacionados con Twitter	39
2.2.4	Datos sobre Twitter.....	39
2.3	TreeTagger	41
2.3.1	Modelo de Markov.....	41
2.3.2	¿Qué es TreeTagger?.....	44
2.3.3	Part of Speech	48
2.3.4	TreeTagger vs. Connexor	50
2.4	Google y Bing.....	51
2.4.1	Finalidad	51
3	ANÁLISIS.....	55
3.1	Visión general del sistema	55



3.2	Requisitos Funcionales	55
3.3	Requisitos No Funcionales.....	58
3.3.1	Interfaces externas.....	58
3.3.2	Interfaz de usuario	58
3.3.3	Requisitos de rendimiento.....	59
3.4	Casos de usuario	60
3.4.1	Descripción gráfica.....	60
3.4.2	Descripción Textual.....	61
4	DISEÑO	67
4.1	Funcionamiento del sistema.....	67
4.1.1	Fase 1 – Inicio.....	69
4.1.2	Fase 2 – Recuperación Tweets.....	69
4.1.3	Fase 3 – Identificación de adjetivos	70
4.1.4	Fase 4 – Análisis de adjetivos.....	71
4.1.5	Fase 5 – Sentimiento de la Marca.....	73
4.2	Algoritmo de Obtención de Connotación del Adjetivo	73
4.3	Prototipo de la interfaz	77
4.4	ARQUITECTURA DEL SISTEMA	¡Error! Marcador no definido.
4.4.1	Arquitectura de 3 capas	80
4.4.2	Diagrama de clases	82
4.4.3	Diagrama de secuencia	85
5	IMPLEMENTACIÓN Y DESARROLLO.....	93
5.1	Tecnologías utilizadas.....	93
5.2	Interfaz	93
6	GESTIÓN DEL PROYECTO.....	99
6.1	Modelo del ciclo de vida del software.....	99
6.2	Planificación	100
6.3	Presupuesto	104
6.3.1	Oferta	104
6.3.2	Periodo de validez.....	104
6.3.3	Costes.....	104
6.3.4	Costes Directos Totales	107
6.3.5	Costes Indirectos	107
6.3.6	Beneficio	107



6.3.7	Margen de Riesgo.....	108
6.3.8	Presupuesto Total.....	108
6.3.9	Forma de pago.....	108
6.3.10	Garantía.....	110
7	CONCLUSIONES.....	114
8	TRABAJO FUTURO.....	118
9	Bibliografía.....	122
	Anexo I. GLOSARIO DE TÉRMINOS.....	126
	Anexo II. MANUAL DEL USUARIO.....	130





Índice de Tablas

Tabla 1.	Aplicaciones del PNL	36
Tabla 2.	Minería de opiniones	37
Tabla 3.	País – Número de usuarios	40
Tabla 4.	Resumen clases gramaticales	49
Tabla 5.	Formato de la tabla	56
Tabla 6.	RF 001 – Recibir marca	56
Tabla 7.	RF 002 – Buscar tweets en función de la marca	56
Tabla 8.	RF 003 – Identificar adjetivos	57
Tabla 9.	RF 004 – Valorar tweets	57
Tabla 10.	RF 005 – Clasificar adjetivos	57
Tabla 11.	RF 006 – Valorar marca	57
Tabla 12.	RF 007 – Mostrar resultado	57
Tabla 13.	RNFIE 001 – Buscador de Tweets	58
Tabla 14.	RNFIE 002 –Api de Google	58
Tabla 15.	RNFIE 003 – Librería de treeTagger	58
Tabla 16.	RNFIU 001 – Todos los sistemas operativos.....	58
Tabla 17.	RNFIU 002 – TreeTagger	59
Tabla 18.	RNFIU 003 – Accesibilidad para personas mayores	59
Tabla 19.	RNFIU 004 – Accesibilidad para usuarios principiantes	59
Tabla 20.	RNFIU 006 – Conexión a internet.....	59
Tabla 21.	RNFIU 007 – Portátiles y PC	59
Tabla 22.	RNR 001 – Número máximo de tweets	59
Tabla 23.	RNR 001 – Número máximo de tweets	60
Tabla 24.	Formato de la tabla de requisitos	61
Tabla 25.	CU 001 – Pedir valoración.....	62
Tabla 26.	CU 002 – Ver Resultados	62
Tabla 27.	CU 003 – Tweet.....	62
Tabla 28.	CU 005 – Global	63
Tabla 29.	CU 004 – Por sentimiento.....	63
Tabla 30.	Estructuras básicas	71
Tabla 31.	Adjetivos comparativos.....	72
Tabla 32.	Superlativos absolutos.....	72
Tabla 33.	Planificación del Proyecto.....	101
Tabla 34.	Coste por Hora/Categoría.....	105
Tabla 35.	Total Coste Personal	105
Tabla 36.	Total Coste de equipos.....	105
Tabla 37.	Total Coste de software.....	106
Tabla 38.	Total Coste Material fungible	106
Tabla 39.	Total Coste de viajes y dietas.....	106



Análisis de sentimientos acerca de marcas mediante Tweets

Tabla 40.	Total Costes directos.....	107
Tabla 41.	Total Costes indirectos	107
Tabla 42.	Beneficio.....	107
Tabla 43.	Margen de riesgo.....	108
Tabla 44.	Presupuesto total sin IVA	108
Tabla 45.	Presupuesto total con IVA	108
Tabla 46.	Resumen de Forma de Pago.....	109



Índice de Ilustraciones

Ilustración 1.	Página principal de Facebook	21
Ilustración 2.	Página principal de Facebook	22
Ilustración 3.	Página principal de Facebook	22
Ilustración 4.	Página principal de Tuenti	23
Ilustración 5.	Página principal de Facebook	23
Ilustración 6.	Página principal de Twitter	24
Ilustración 7.	Árbol de análisis	36
Ilustración 8.	Extensión demográfica	40
Ilustración 9.	Red bayesiana dinámica simple y modelo de Markov....	43
Ilustración 10.	Árbol de decisión algoritmo ID3.....	45
Ilustración 11.	Resultados de las pruebas.....	46
Ilustración 12.	Recursividad del árbol.....	46
Ilustración 13.	Frecuencia	47
Ilustración 14.	Árbol de 3 niveles.....	47
Ilustración 15.	Probabilidades de éxito	48
Ilustración 16.	Diagrama categorías lingüísticas inglesas	49
Ilustración 17.	Casos de uso	61
Ilustración 18.	Funcionamiento del sistema.....	68
Ilustración 19.	Funcionamiento del sistema - Fase 1.....	69
Ilustración 20.	Funcionamiento del sistema - Fase 1.....	69
Ilustración 21.	Funcionamiento del sistema - Fase 3.....	70
Ilustración 22.	Funcionamiento del sistema - Fase 4.....	71
Ilustración 23.	Funcionamiento del sistema - Fase 5.....	73
Ilustración 24.	Explicación de los buscadores Google y Bing.....	74
Ilustración 25.	Resultado de la comparación sucio con guapo – feo ...	75
Ilustración 26.	Resultado de la comparación sucio con calor – frío	75
Ilustración 27.	Resultado de la comparación sucio con bien – mal	76
Ilustración 28.	Resultado de la comparación sucio con lleno – vacío ..	76
Ilustración 29.	Prototipo de la pantalla inicial de la aplicación	78
Ilustración 30.	Prototipo de la pantalla inicial después de la búsqueda	79
Ilustración 31.	Modelo arquitectura de 3 capas	80
Ilustración 32.	Modelo arquitectura de 3 capas para el proyecto	81
Ilustración 33.	Diagrama de clases.....	82
Ilustración 34.	Diagrama de clases – Google	83
Ilustración 35.	Diagrama de clases – Twitter	84
Ilustración 36.	Diagrama de clases – Análisis de sentimientos acerca de marcas mediante Tweets	85
Ilustración 37.	Diagrama de secuencia.....	86



Ilustración 38. Diagrama de secuencia. Twitter – Análisis de sentimientos acerca de marcas mediante Tweets.....	87
Ilustración 39. Diagrama de secuencia. Análisis de sentimientos acerca de marcas mediante Tweets - TreeTagger	88
Ilustración 40. Diagrama de secuencia. TreeTagger - Google	90
Ilustración 41. Interfaz de la aplicación antes de iniciar una búsqueda.....	94
Ilustración 42. Interfaz de la aplicación al introducir una marca	95
Ilustración 43. Interfaz de la aplicación después de iniciar una búsqueda.....	96
Ilustración 44. Ciclo de vida en cascada	99
Ilustración 45. Diagrama de Gantt	103
Ilustración 46. Primera pantalla	131
Ilustración 47. Segunda pantalla	132
Ilustración 48. Tercera pantalla	133



CAPÍTULO 1

Introducción

Este capítulo se va a utilizar para introducir al usuario dentro del tema que nos atañe, por lo que se presentarán diversos conceptos relacionados con tecnología, Internet y redes sociales.

Por otro lado se pasará a explicar el problema que se ha planteado, algunas soluciones y alternativas barajadas a dicho problema, y la solución final que se ha utilizado para desarrollar el proyecto.

A continuación se pasará a detallar los objetivos que debe cumplir el proyecto. Este paso es importante puesto que se tiene que asegurar que el proyecto va a ser desarrollado y generado de una forma satisfactoria, tanto para el cliente, como para el proveedor y así poder obtener una aplicación eficiente y eficaz.

Para concluir con este capítulo, se realizará un pequeño resumen de los apartados por los que está compuesto este documento.



1 INTRODUCCIÓN

1.1 Contexto

Todo lo relacionado con la tecnología, la informática, el Internet o saber cómo se utiliza un ordenador, son cuestiones que cada día va cogiendo mayor auge dentro de nuestras vidas, por lo que hay que saber convivir con ello.

Hace unos años, estas áreas no estaban muy desarrolladas pero hoy en día están dentro de nuestras vidas, tan dentro que incluso una cosa tan cotidiana, como es ir a hacer la compra, hoy en día se puede hacer a través de Internet mediante un Smartphone, un ordenador, una tablet o incluso una televisión.

Otro ejemplo puede ser la “*caja amiga*”, una caja en la que el comprador se auto-cobra los productos, sin necesidad de tener una cajera que le atienda. De igual forma, si vamos a un centro comercial es posible que hagamos uso de un ordenador para conocer si determinado establecimiento se encuentra o no en el centro.

Estos son solo algunos pequeños ejemplos de cosas que hoy en día se pueden hacer gracias a lo que ha avanzado la tecnología en el campo de la informática.

La tecnología ha tenido un impacto tan grande en nuestras vidas y el concepto de red social ha llegado a calar tan hondo, que los usuarios tienen las redes sociales tan interiorizadas que en muchos casos no pueden vivir sin ellas.

Las redes sociales, son plataformas que requieren de internet para funcionar; están alojadas en servidores y albergan a todo tipo de usuarios.

Estos usuarios se pueden conocer o no, pero cada uno muestra sus inquietudes haciendo que el resto de usuarios puedan discutir, comentar o compartir lo que ha mencionado otro usuario o en este caso amigo.

La tecnología ha llegado tan dentro de nuestras vidas, que hay incluso ciertos usuarios de redes sociales que utilizan estas plataformas como “diario personal”.

Al hacer referencia a “diario personal” se quiere decir que ponen a disposición de todos sus amigos en las redes sociales: lo que les ocurre, lo que piensan, lo que hacen o donde están. Esto, si no se sabe llevar con cierta prudencia, puede llegar a producir muchos problemas, aunque esta cuestión no se abordará como parte de este proyecto concreto.



En España, las redes sociales de mayor importancia en cuanto a número de usuarios registrados son: *Facebook*, *Tuenti* y *Twitter*. Si bien todas ellas son parecidas en el sentido de que cumplen la premisa de concentrar a gente para tratar diversos temas de conversación, cada una de ellas tiene su estilo propio.

En las siguientes líneas se pasará a comentar, un poco, cada una de las redes sociales para dar al lector una idea global sobre el tema que se va a tratar:

- En **primer lugar** se explicará un poco la finalidad de *Facebook*, como surgió y para que se utilice hoy en día.
- En **segundo lugar** se comentarán las características de *Tuenti*, la cual tiene mucho parecido con Facebook.
- Por **último**, vamos desglosar *Twitter*, que es la más joven de las tres, además será la red social en la que se basa la *app* en la que desarrollaremos nuestro proyecto.

1.2 Facebook



Ilustración 1. *Página principal de Facebook*

Fue una de las primeras redes sociales que surgió, fue creada en San Francisco, Estados Unidos, originalmente era para gente que perteneciera a la universidad de Harvard, pero hoy en día es abierta y cualquiera que tenga un correo electrónico puede pasar a formar parte ella.

Facebook, ofrece una amplia oferta de servicios y *app's* para poder realizar acciones como: “*postear*” algo en tu muro y que alguien pueda comentar ese “*post*”, subir una foto y etiquetar a la gente, que aparezca en la misma y poder comentarla o darle un “*me gusta*”.

Si alguien escribe algo y te gusta, a continuación hay un botón (“*Me Gusta*”), el cual va contabilizando a cuanta gente le ha gustado ese “*post*” o foto.

Facebook, no solo te da la oportunidad de poder conversar y mostrar tu vida a través de internet, si no, que también te propone juegos, para que puedas jugar con tus amigos *online*. El juego más conocido es el de *FarmVille*, llegó a tener tanto éxito que se tuvo que crear una *app* para *smartphones* para que la gente pudiera jugar a través de ellos.



Aparte de las diversas funcionalidades que te da *Facebook*, como pueden ser aplicaciones o juegos, su verdadero atractivo es el de poder sociabilizar en red. Para sociabilizar en red se tienen diversas funciones como pueden ser: aceptar, bloquear o solicitar amistad.

Por otro lado tienes privacidad para que el resto de usuarios puedan o no ver tus fotos o puedan leer o no los post que aparezcan en tu muro. Si tus amigos no están en la plataforma no podrás interactuar con ellos, ni etiquetarles en fotos.

En la **Ilustración 2** se muestra la página principal a la que se llega cuando se introduce www.facebook.com en cualquier explorador.

The screenshot shows the Facebook homepage layout. At the top, there's a blue header with the Facebook logo on the left and login fields on the right. The login fields include 'Correo electrónico o teléfono' and 'Contraseña', with an 'Entrar' button. Below these are checkboxes for 'No cerrar sesión' and a link for '¿Has olvidado tu contraseña?'. The main content area is divided into two sections. On the left, 'Tu biografía de Facebook' features a video player showing a person in a pool. On the right, the 'Regístrate' section includes a registration form with fields for 'Nombre', 'Apellidos', 'Tu correo electrónico', 'Vuelve a escribir tu correo electrónico', 'Nueva contraseña', 'Sexo' (a dropdown menu), and 'Fecha de nacimiento' (with separate dropdowns for 'Día', 'Mes', and 'Año'). A green 'Regístrate' button is at the bottom of the form. A small disclaimer about terms and conditions is visible below the date field.

Ilustración 2. Página principal de Facebook

1.3 tuenti



Ilustración 3. Página principal de Facebook

Se creó después de Facebook, es una comunidad privada, por lo que para poder entrar se necesita una cuenta de correo y un amigo que ya esté dentro.

Es muy parecido a Facebook, puesto que su finalidad es la misma, reunir a gente para que suban fotos, las comenten, comenten en sus muros; vamos para que cada uno exprese sus ideas de la forma que quiera.



Para hacernos una idea del perfil del usuario al que está enfocada esta red social, solo hay que fijarse en su nombre; *Tuenti* tiene una gran similitud a la palabra inglesa *twenty* (veinte) por lo que los usuarios, que van a utilizar esta aplicación, tendrán alrededor de esta edad.

Por esa razón la edad mínima para acceder a dicha red social son quince años, pero como es bien sabido por todos y debido a la popularidad de la misma hay muchos usuarios que no alcanzan la edad mínima y falsean sus datos para poder darse de alta.

En el ámbito legal existe una gran diferencia, a nuestro criterio, entre *Facebook* y *Tuenti*. *Facebook* se rige por la legislación californiana y sin embargo sobre *Tuenti* exige legislación española. Bajo nuestro punto de vista las leyes españolas protegen la privacidad de la información mucho mejor que las leyes californianas.

En la **Ilustración 4** se muestra la página principal a la que se llega cuando se introduce www.tuenti.com en cualquier explorador.

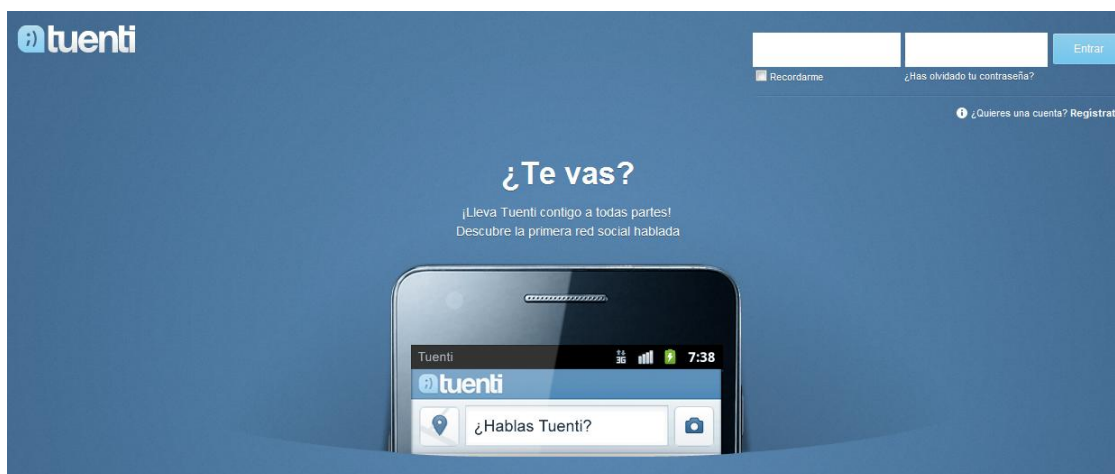


Ilustración 4. Página principal de Tuenti

1.4 twitter



Ilustración 5. Página principal de Facebook

Es en la red social en la que se va a centrar este proyecto, por lo que se va a comentar brevemente sobre ella puesto que dentro del apartado **2** se van a dar más detalles sobre ella, concretamente en el punto **2.2**.

Apareció junto con *tuenti* en el 2006, pero en este caso *twitter* se desmarca del resto gracias a las características particulares que tiene.

La característica más representativa frente a *tuenti* y *facebook* es que *twitter*, solo se tienen **140 caracteres** para escribir, no se pueden sobrepasar, pero se pueden enviar tantos mensajes como se quiera y se utiliza fundamentalmente para dar opiniones, comentar algo sobre noticias, fotos o videos, los cuales se pueden incluir en el mismo *tweet*.

1.4.1 Dispositivos donde puede utilizarse

Twitter puede ser utilizado bien desde un navegador web, accediendo directamente a la propia web de *twitter*, o bien a través de aplicaciones desarrolladas específicamente para los distintos modelos de *smartphones*, *tables* y dispositivos móviles existentes en el mercado.

En la **Ilustración 6** se muestra la página principal a la que se llega cuando se introduce www.twitter.com en cualquier explorador.



Ilustración 6. Página principal de Twitter

1.5 El reto

Este proyecto surge, por una pregunta que se plantea un *community manager* de una empresa: “¿Qué pensará la gente, que escribe en Internet, sobre mi marca?”.

Aunque parece una pregunta sencilla no es fácil de responder. Para empezar, al ser Internet una fuente de información tan amplia, se tiene que poder hacer un poco más pequeño el campo de búsqueda.



En este caso, tal y como se comentó en el [1.1](#), el proyecto se centrará en dar respuesta a esa pregunta para el caso específico de la red social *Twitter*.

Después de investigar con la herramienta, *Twitter*, y conocer su funcionamiento, pasamos a plantearnos el siguiente reto: ***“De qué forma se pueden conseguir todos los tweets que aparecen en twitter de una forma fácil y rápida.”***

Hasta aquí puede ser una cosa trivial, puesto que en tu muro de *twitter* te aparecen todos los *tweets* de una forma fácil y rápida. En este momento es en el que llegamos al RETO, los *tweets* que se quieren considerar u obtener tienen algunas puntualizaciones, no vale cualquier *tweet*. Las características son las siguientes:

- Tiene que estar escrito en español.
- Contener dentro de su texto, el nombre de la marca especificada.
- Obtener una valoración (positiva, negativa o neutra) sobre el *tweet*.

Aún bajo estas premisas conseguir dar respuesta a la pregunta inicial, sigue siendo algo complejo ya que, como se puede leer en el punto [2.2.2](#), esta red social tiene millones de usuarios repartidos por todo el mundo. Por ello, y para poder asumir el reto, hay que comprobar si *Twitter* puede proporcionar directamente, las dos características que se necesitan para llevar a cabo el proyecto.

A continuación se mostrarán las dos características y se explicará si se puede obtener directamente o se tiene que hacer algún cambio.

Tweets escritos únicamente en español. Esta parte no está implementada en la interfaz de *twitter*, el usuario no puede elegir si se pueden buscar unos *tweets* con unas características concretas como puedes ser: por idioma, por lo que en este momento se empieza a complicar algo más el reto, puesto que ya hay que entrar dentro de su API y utilizar alguna función que nos devuelva lo que estábamos buscando. Por lo que esta parte tiene que ser implementada para el proyecto.

Conseguir SOLO los *tweets* que contengan el nombre de la marca. Ésta característica, *twitter* nos la puede dar puesto que la propia red social tiene un buscador integrado, por lo que solo introduciendo el nombre de la marca ya se podrían encontrar todos los *tweets*, pero en este caso buscaría todos, incluyendo aquellos que no han sido escritos en castellano. Por tanto el buscador no es todo lo eficiente que necesitaríamos puesto que nos da información extra que no se necesita.



Valoración del *tweet*. La red social no da ni de forma directa ni indirecta una valoración del *tweet*. La clasificación que se está buscando es en función de los adjetivos que tenga el *tweet* dar un valor *positivo*, *negativo* o *neutro*. Esta parte es la más complicada, porque no es una opción contemplada dentro del API de *twitter*, por lo que se tiene que desarrollar por completo.

Al haber analizado las características de nuestro producto y ver que la red social en la que se basa el proyecto no es capaz de darlo de forma directa, se puede afirmar, que el proyecto que se está desarrollando en este documento es viable.

1.6 Soluciones propuestas

Como ya se conoce el reto, al que hay que enfrentarse, se ha estado pensando en las posibles soluciones, que se podrían llevar a cabo para solucionarlo.

Por lo que este es el punto, es en el que se van a mostrar las soluciones, que se han planteado para poder dar solución al reto. En este caso se planteará la solución y si puede ser viable o no, junto con sus pros y contras.

Primera solución coger todos los *tweets* que van saliendo y mirar si en ellos aparece la marca en concreto.

Esto no es ni fácil, ni rápido, puesto que se necesitaría a una única persona “full time”, día y noche para poder comprobar eso y sería prácticamente imposible.

Segunda solución es la creación de un *hashtag* único con el nombre de la marca, para que el usuario, que quisiera hablar sobre ella la pusiera en su *tweet*.

Esta solución ya es más viable, fácil y rápida, pero no es muy óptima, puesto que no todo el mundo tiene porqué saber, que para esa marca ha creado un *hashtag*, el cual se tiene que utilizarlo cada vez que se quiera hablar de ella.

Por esta razón, esta solución no es válida para resolver el reto, puesto que en muchas ocasiones se perderían un gran número de *tweets*.

Tercera solución es la creación de una aplicación, la cual se va a unir internamente con *twitter*, pero van a ser aplicaciones independientes cara al usuario.



La aplicación se encargará de buscar una marca introducida por el usuario y aparecerán directamente todos los tweets relacionados con ella.

Esta **tercera solución** es la que **se va a llevar a cabo**, puesto que es la que cumple las premisas del proyecto.

La aplicación que se creara se llamaría “**Análisis de sentimientos acerca de marcas mediante Tweets**”, el manager solo tendrá que introducir la marca de su empresa y activar el buscador, pasado unos segundos, se mostrarán todos los *tweets* relacionados con la marca y ya estarán evaluados y señalados con sus correspondientes “sentimientos”: positivo, negativo o neutro.

Incluso para que sea más rápido se le indicará en la parte superior de la *app* con un diagrama muy sencillo que dará la valoración global de los *tweets*.

1.7 Objetivos

Una vez detectada y planteada la problemática a la que nos enfrentamos, se tienen que establecer los objetivos que se van a llevar a cabo para que el proyecto cumpla las condiciones iniciales.

1.7.1 Consideraciones previas

Este proyecto surge por la necesidad que tienen las empresas en saber, la opinión que tienen sus clientes sobre sus productos. Por tanto utilizando la técnica de análisis de sentimientos se puede obtener información subjetiva del texto.

Por la necesidad comentada anteriormente, se ha recopilado información sobre el tema concluyendo que es una técnica muy utilizada y cada vez con mayor auge.

No solo se ha encontrado información, también se han encontrado diversas herramientas informáticas que se utilizan para determinar, el sentimiento que ha plasmado un usuario en un mensaje de texto en la red.

Este sentimiento se consigue a través de ciertas palabras que aparecen en el comentario que ha dejado el usuario. Todo el análisis se tiene que hacer en conjunto, coger todo el texto y saber cómo se ha escrito, aunque se fije la atención en algunas palabras que puedan tener una mayor connotación.

A pesar de encontrar diversas herramientas informáticas, se puede concluir que ninguna se parece a la que se expone para este proyecto. Como ya se ha comentado, esta aplicación va a ser un complemento para un *community manager*, la cual va a ser utilizada en tiempo real sincronizándose con la red social *twitter*.



Por tanto se simplifica la herramienta y la personaliza para esta red social, este caso es muy concreto aunque su funcionamiento interno es totalmente portable a cualquier otra red social o necesidad del cliente.

En este caso se ha decidido fijarse en *twitter* puesto que es una red social en auge, con la que utilizando solo 140 caracteres se puede expresar lo que uno quiera y las respuestas a ciertos impulsos son rápidamente *twitteadas*.

1.7.2 Objetivos

Una vez presentadas las consideraciones previas, se tiene que presentar una aplicación diferente y personalizada pero que a su vez, pueda ser portable, pudiendo ser utilizada en cualquier otra red social o plataforma.

Por lo que este proyecto va a tener las cualidades mostradas anteriormente, siendo una herramienta de ayuda esencial para el *community manager*, pudiendo mostrarle qué opinión tienen los usuarios sobre una marca en concreto dentro de *twitter*.

- El **análisis de sentimiento**, es lo más importante de este proyecto y por tanto es a lo que más hay que prestarle atención. En los capítulos siguientes se mostrará la importancia del análisis de sentimientos y como se ha llevado a cabo y que cuestiones se han tenido en cuenta para poder analizar los diferentes *tweets*.

1.8 Estructura del documento

La información que se recoge en el presente documento se va a estructurar en diversos capítulos, los cuales son:

- **Introducción:** Introducción al usuario dentro del tema que nos atañe, por lo que se comentarán temas relacionados con tecnología, internet y redes sociales. Se definirán las diferentes redes sociales y sus características.

Una vez introducidos y explicados estos conceptos se pasará a explicar el problema que se nos ha planteado, una vez explicado el problema, se pasarán a explicar algunas soluciones que se han planteado hasta dar con la idónea que lo solucionara.

A continuación pasaremos a detallar los objetivos que debe cumplir el proyecto y que de esa forma sea satisfactorio tanto para el cliente como para el proveedor.



Por último, para acabar con este capítulo, se realizará un pequeño resumen de los apartados por los que está compuesto este documento.

- **Estado del arte:** El objetivo de este capítulo es aproximar al lector a los términos de análisis de sentimientos y *twitter*, y a la aplicación de *treeTagger*.

El primer concepto, análisis de sentimientos, es la parte fundamental de este proyecto, puesto que todo lo que aquí se defina va a ir relacionado con este concepto.

El segundo concepto es *twitter*, como ya se comentó, es la red social con la que se va a trabajar. Esta red social se unirá con la aplicación, la cual se está desarrollando en este proyecto, pero para el usuario van a ser dos cosas independientes.

El último concepto es *treeTagger* es un programa que se utiliza para saber la morfología de la palabra que se introduzca en el mismo. Con este programa se podrá conocer el “sentimiento” de los *tweets*, puesto que podremos coger cada palabra por separado y saber qué tipo de palabra es.

- **Análisis:** Se va a explicar toda la aplicación de una forma más técnica, a través de los requisitos software y hardware y casos de uso.
- **Diseño:** Se mostrará en qué tecnologías y en qué contexto se puede utilizar, junto con el prototipo inicial, el cual servirá de base para la interfaz final.
- **Implementación y desarrollo:** Se mostrará la aplicación final que se entregaría al cliente, pero como ya se ha comentado a lo largo del documento, es más importante la investigación sobre el análisis de sentimientos más que la funcionalidad de la aplicación. A pesar de haber creado un prototipo y un diseño final.
- **Gestión del proyecto:** Se mostrará todo lo relacionado con el ciclo de vida que se ha utilizado, la planificación diseñada, los detalles económicos correspondientes a la presente oferta, así mismo como la forma de pago; definiendo los términos de la garantía, los cuales ofrece la aplicación.
- **Conclusiones:** Se presentarán las consideraciones a las que se han llegado una vez concluido el proyecto.



Análisis de sentimientos acerca de marcas mediante Tweets

- **Trabajo Futuro:** Se comentarán en que líneas de trabajo se puede desarrollar o mejorar la aplicación en un futuro.
- **Referencias:** Páginas Web, artículos o libros en los que se han buscado información para poder realizar el presente documento.



CAPÍTULO 2

Estado del Arte

Este capítulo se va a utilizar para realizar una aproximación a: *Twitter*, *Análisis de Sentimientos*, *TreeTagger* y *Google y Bing*.

Estos conceptos son muy importantes, puesto que son la base del proyecto. Se tienen que tener muy claros y saber perfectamente para que sirve cada uno de ellos y cuál es su función.

Twitter, será la fuente de la que se sacarán los tweets que posteriormente se analizarán.

Análisis de sentimientos, es la parte fundamental de este proyecto. Se explicará cómo se puede conocer el pensamiento de los usuarios, a través de las palabras utilizadas en sus comentarios.

TreeTagger, programa que se utiliza para dividir y clasificar cada palabra del *tweet*.

Google y Bing, ayudarán a mejorar el análisis de sentimientos.



2 ESTADO DEL ARTE

2.1 Análisis de sentimientos

Es la parte más importante de este proyecto, por tanto es en la que ocupará más páginas en este proyecto. Se quiere explicar bien los conceptos e intentar hacer ver que realmente es un proceso muy utilizado y que cada vez es más demandado.



2.1.1 Definición

El término análisis de sentimientos ha sido definido de distintas formas. Por ejemplo:

- *“El análisis de sentimiento es una rama del procesamiento del lenguaje natural, que determina la actitud de un autor respecto a un tema. La actitud puede ser la opinión, el estado afectivo o el estado que intenta comunicar.” (Sentisis, 2013)*
- *“El análisis de sentimientos se utiliza para clasificar los documentos según la opinión que expresa su autor. Esta nueva área que combina PLN y minería de textos, incluye una gran cantidad de tareas, que han sido tratadas en mayor o menor medida. Existen principalmente dos formas distintas de enfrentarse a este problema: aplicando aprendizaje automático o aplicando un enfoque semántico.*

El análisis de sentimientos suele ser utilizado principalmente para determinar la polaridad de las opiniones a nivel de documento, frase o característica, y para determinar si un documento contiene opiniones. Existen muchos trabajos en el campo del análisis de sentimientos, habiéndose aplicado en multitud de dominios, pero la mayor parte de ellos han sido realizados sobre campos de documentos en el idioma inglés.



La investigación en el análisis de sentimientos en el lenguaje español es reducida, pero se puede destacar la que está llevando a cabo por el grupo ITALICA de la universidad de Sevilla. El grupo SINAI de la universidad de Jaén también lleva algún tiempo trabajando en esta



temática pero centrándose hasta hace unos meses solamente en la lengua árabe.” (Ureña & Martínez, 2011)

Estrechamente relacionados con el análisis de sentimientos se encuentran por tanto el *Procesamiento del Lenguaje Natural* (PLN) y la *Minería de opiniones*.

2.1.1.1 Procesamiento del Lenguaje Natural (PLN)

El procesamiento de lenguaje natural es una disciplina dentro de la inteligencia artificial.

No es una disciplina de nueva aparición, las primeras investigaciones se remontan al origen de los primeros ordenadores, aunque para la parte que nos atañe, hoy en día es cuando más investigaciones se están realizando sobre este tema y más concretamente todo lo relacionado con el análisis de textos.

Por otro lado el procesamiento de lenguaje natural no solo se utiliza para el análisis morfológico o sintáctico de palabras o frases, sino que también se utiliza para dotar de aparente “*inteligencia*” a las máquinas.

Para poder explicar con mayor claridad este punto se pondrá un ejemplo: La primera premisa es, tener un texto dentro del ordenador y se requiere sacar cierta información del mismo a través de ciertas preguntas. Esto solo se puede ser viable siempre y cuando el procesamiento del lenguaje natural se encuentre por detrás del proceso.

Por tanto, gracias a este sistema, se puede ejecutar un código que la máquina lo procese y los usuarios puedan recibir las respuestas a las preguntas planteadas, por este motivo se dice que se está dotando de “*inteligencia*” a la máquina. Por otro lado, los buscadores, a grosso modo, es lo que realizan para poder dar resultados sobre las preguntas que se hacen.

La gran mayoría de sistemas, los cuales tienen o involucran en su funcionamiento el proceso de lenguaje natural, se basan en la siguiente estructura, la cual consta de 3 pasos:

- El **primer paso**; Elegir el texto a tratar, como no se puede utilizar directamente de la forma obtenida, se le tiene que aplicar un cambio, para que pueda ser procesado.

El cambio, que se tiene que realizar, consiste en la transformación del texto en lenguaje natural a una representación formal. Para que todo siga correcto, las características que estuvieran en el texto en lenguaje natural tienen que mantenerse en la transformación, para no perder nada de información.



- El **segundo paso**; consiste en procesar la representación formal del texto escogido para obtener el resultado deseado.
- Por **último**; este paso no es siempre necesario, se presentará el resultado del texto procesado y si es necesario, ésta es la parte opcional, se volverá a transformar a lenguaje natural.

Una vez explicado todo lo que es el procesamiento del lenguaje natural y en que se basa, hay que tener en cuenta un “*Hándicap*” a superar, **la ambigüedad**, ésta característica se da a diferentes niveles gramaticales:

- **A nivel léxico**, muchas veces se pueden encontrar palabras, que en función del contexto en el que se utilicen, pueden dar un sentido u otro a la frase. Por lo que se tiene que deducir el concepto a partir del concepto oracional o el conocimiento básico. Gracias a diccionarios gramaticales, bases de conocimiento y correlaciones estadísticas ya se está superando este nivel.
- **A nivel referencial**, en lengua escrita, al menos en español, se puede sustituir sujetos por pronombres, por lo que se pueden encontrar dificultades a la hora de saber si es un pronombre o un determinante.
- **A nivel estructural**, en este nivel se tiene que tener cuidado puesto que hay frases que se pueden escribir de dos formas diferentes, utilizando las mismas palabras, pero según el orden en el que se dispongan, pueden significar cosas diferentes.
- **A nivel pragmático**, en este nivel ya entra en juego: los dobles sentidos, la ironía y los juegos de palabras, puesto que estas características pueden hacer que una frase cambie totalmente de significado y por tanto se pueda mal interpretar.

Este “*Hándicap*” se está tratando de solucionar a través de los árboles de análisis, que se encargan de realizar una traducción del lenguaje natural a una representación interna, sin ambigüedad, pudiendo analizar el texto sin perder la esencia del mismo y así superar el “*Hándicap*”.

En la **Ilustración 7** se muestra una representación de un árbol de análisis. Como se puede observar el árbol de análisis muestra un grafo, en el que se representan las relaciones jerárquicas que existen entre los constituyentes sintácticos (palabras de la oración).

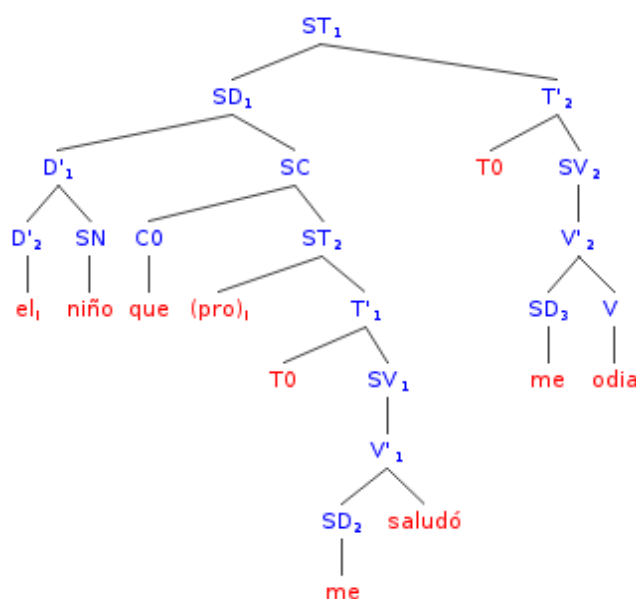


Ilustración 7. Árbol de análisis

Como se ha mencionado, el procesamiento del lenguaje natural se utiliza para diversos motivos y con diferentes funcionalidades, a continuación se va a mostrar en la **Tabla 1**, a modo de resumen, todas las aplicaciones en las que se puede utilizar.

Aplicaciones	
Síntesis del discurso	Generación de lenguajes naturales
Análisis del lenguaje	Traducción automática
Comprensión del lenguaje	Respuestas a preguntas
Reconocimiento del habla	Recuperación de la información
Síntesis de voz	Extracción de información

Tabla 1. Aplicaciones del PNL

2.1.1.2 Minería de opiniones

Se suele denominar minería de opiniones al conjunto de técnicas que se utilizan para el procesamiento del lenguaje natural, lingüística computacional y minería de textos, teniendo como objetivo la extracción de información subjetiva a partir de contenidos generados por un usuario.

Esta definición recoge el propósito del presente proyecto: a partir de un comentario textual ser capaces de extraer un valor tangible y directo, el cual puede ser tanto positivo como negativo.

La minería de opiniones es un campo cada vez más extenso, que está relacionado con el análisis de componentes subjetivos, los cuales están implícitos en los contenidos que generan los usuarios. Las aplicaciones que existen se pueden dividir en dos, las cuales se muestra resumidas en la **Tabla 2**:



Tarea	Descripción
Detección de polaridad	Saber diferenciar si la opinión que se ha creado es positiva o negativa, esta es la utilidad básica. Esta utilidad siempre se puede complicar, otra opción es la de crear una escala numérica y asociarla a la opinión creada obteniendo un ratio objetivo.
Análisis del sentimiento basado en características	Esta tarea es mucho más compleja que la anterior, puesto que en esta lo que se quiere realizar, es coger cada característica del producto, las cuales aparece en la opinión del usuario y para cada una de ellas extraer la polaridad.

Tabla 2. Minería de opiniones

Durante la búsqueda de información sobre la minería de opiniones se ha encontrado dos sitios web muy interesantes: el primer sitio web es (Monsalve, 2010) y el segundo es (Fernández, Boldrini, Gómez, & Martínez-Barco, 2010).

En el primer sitio web se realiza un estudio sobre el análisis de sentimientos y la minería de opiniones. Se destaca la importancia que está surgiendo en la aplicación de estas técnicas; pudiendo relacionarlas en los próximos años con: análisis previos a campañas de publicidad o estudios de mercado para un producto.

Por otro lado, cuenta que este tipo de servicios supondrían una gran ayuda para ciertos departamentos o personas dentro de las empresas, con este tipo de aplicaciones se ayudaría a esas personas a tomar decisiones sobre: campañas de publicidad, productos que sacar al mercado o mejorar ciertas características del producto, entre una infinitud de opciones.

Este es otro punto que comparte con este proyecto puesto, puesto que está pensado para ayudar a los *community managers* a que puedan realizar su trabajo.

Por tanto en un par de años seremos capaces de saber los sentimientos de una persona simplemente, con las palabras que utiliza dentro de una opinión. (Monsalve, 2010)

El segundo sitio web se llama Emotiblog, se encarga de ir recopilando entradas de blogs; en estas entradas se van realizando diferentes algoritmos para detectar expresiones subjetivas que aparecen en ellas, en este caso se está abriendo el mercado para los nuevos géneros textuales que están apareciendo en la Web 2.0, para facilitar y detectar este tipo de información de una forma más rápida en las entradas de blogs, se han utilizado sistemas de aprendizaje automático.

En el artículo comenta que esas firmas de aprendizaje automático no son suficientes y se han ido introduciendo otros tipos de técnicas como son el análisis de sentimientos y la minería de opiniones. (Fernández, Boldrini, Gómez, & Martínez-Barco, 2010)



2.1.2 ¿Para qué se utiliza?

Esta pregunta se responde con la finalidad por la que se realiza este proyecto: *“Saber qué opinión tiene la gente sobre un producto, simplemente con las palabras que utilizan para describirlo.”*

No solo se utiliza con este propósito, pero en este caso, la aplicación que se está desarrollando en este documento, realiza un análisis de las opiniones que se realizan sobre una marca o un producto de esa marca.

A pesar de ser textos cargados de subjetividad, se necesita centrar la atención, en este caso los adjetivos, para poder reflejar y analizar con exactitud lo que el usuario quiere mostrar.

En estos textos se pueden encontrar ironías o dobles sentidos, que se tienen que tener en cuenta para poder analizar con mayor precisión el *tweet*, este asunto se abordó en el punto [2.1.1.1](#)

Por lo que una vez superados todos esos problemas léxicos, se podrá analizar el *tweet* para asignarle “su sentimiento”.

2.2 Twitter

En esta sección se ofrecerá una visión general sobre qué es *Twitter* y sus usos y aplicaciones más comunes.

2.2.1 ¿Qué es Twitter?

Twitter es una red social que se sustenta sobre un servicio de *microblogging*. Es un servicio que permite a los usuarios enviar o publicar mensajes con un máximo de 140 caracteres, llamados “*tweets*”, que generalmente son de texto, aunque en algunas ocasiones pueden introducir links a otros enlaces, en los cuales se pueden encontrar: fotos o noticias de internet.



Las actualizaciones de los mensajes enviados, se van mostrando en la página de perfil del usuario y en tiempo real al resto de los usuarios que hayan elegido recibirlas.

2.2.2 Historia de Twitter

Twitter comenzó en el 2006 con la participación de 18 trabajadores y han tenido que aumentar su plantilla de forma exponencial llegando en el 2012 a ser cuatro veces más y con perspectivas de seguir creciendo.



En sus inicios *Twitter* se financiaba gracias a los anuncios de *addSense* de Google, pero en el momento en el que aumentó de una forma tan significativa el número de usuarios, decidió prescindir de este tipo ingresos y financiarse a través de empresas de capital riesgo.

Posteriormente cambiaron las condiciones de servicio y permitieron a los usuarios introducir publicidad dentro de los *tweets*. Este tipo de publicidad se llama “*tweets* promocionados”, y para evitar que se produjera intrusismo, solo se permite crear un mensaje patrocinado.

2.2.3 Conceptos relacionados con *Twitter*

Para poder conocer bien el mundo de *twitter* hay que tener claras ciertos conceptos que son frecuentemente manejados por sus usuarios.

- ***Tuiteros***: personas que utilizan la aplicación.
- ***Tweet***: es el mensaje corto de 140 caracteres que publica un usuario.
- ***Follower***: es la persona que “sigue” a otra. Ser *follower* conlleva recibir todos los *Tweets*, los cuales genera la persona la cual estás siguiendo. Por otro lado, La popularidad de un usuario se mide por el número de followers que tiene.
- ***Trending topic***: es la palabra que se utiliza para denominar un fenómeno de masas. Un tema puede llegar a ser *trending topic* en el momento en el que un gran número de usuarios comentan sobre él.
- ***Hashtag***: es la palabra que se utiliza para denominar a temas del momento o de moda. La forma de describirlo es: # + nombre, con nombre queremos hacer referencia a cualquier tema o frase o acontecimiento ocurrido.
- ***Retweet***: es la palabra que se utiliza para designar al *tweet* que te han gustado y tú lo vuelves a reenviar a tus *followers*.
- ***Twittear***: el verbo que utilizamos cuando se quiere poner un *tweet*. Cuando alguien pregunta ¿Qué estás haciendo? – *twittear*.

2.2.4 Datos sobre *Twitter*

En este apartado pasaremos a explicar, el impacto social, en cuantos lugares del mundo se encuentra, que tipo de usuarios lo utilizan y de qué forma se utiliza.

El ***impacto social*** que ha generado *twitter* ha sido tan grande que se ha sido empleado como medio de difusión principal de algunas de revoluciones

sociales más importantes de los últimos años, como la egipcia, la tunecina o las protestas electorales en Irán.

Por otro lado, en España se ha utilizado en las distintas campañas electorales, llegando a ser tan relevante que todos los políticos y partidos políticos se han creado cuentas las cuales utilizan o actualizan casi todos los días.

Es importante destacar que *twitter* es altamente participativo y que los propios usuarios pueden contribuir a mejorar la aplicación. Por ejemplo, en 2009 se estrenó la versión en español desarrollada gracias a los propios *tuiteros*, a la que luego siguieron la versión francesa, italiana y por último alemana.

El número de países en los que se utiliza *twitter* es muy extenso, tal y como se muestra en la **Ilustración 8** y en la **Tabla 3** donde se recoge la relación entre países y números de usuarios por cada uno de ellos.

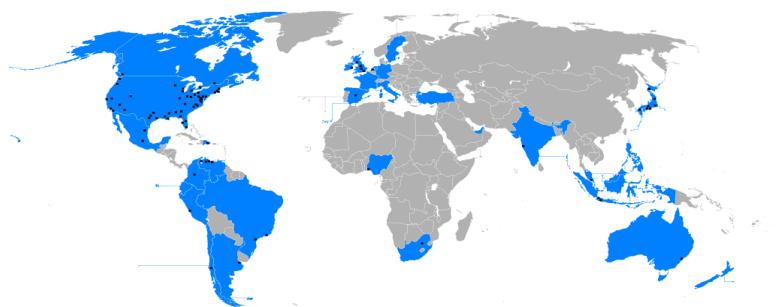


Ilustración 8. Extensión demográfica

País	Usuarios de Twitter
Estados Unidos	107,7 millones
Brasil	33,3 millones
Japón	29,9 millones
Reino Unido	23 millones
Indonesia	19 millones
India	12 millones
México	10,5 millones
Filipinas	8 millones
España	7,9 millones
Canadá	7,5 millones

Tabla 3. País – Número de usuarios

Ha llegado a ser tan relevante la red social, que se han realizado entrevistas a políticos y personajes famosos en tiempo real. Se ha encontrado una página web “<http://www.clasesdeperiodismo.com/2013/01/18/lo-que-debemos-hacer-para-que-una-entrevista-por-twitter-sea-exitosa/>” en la que explican a periodistas como realizar entrevistas a través de *twitter* y que sean exitosas.



2.3 *TreeTagger*

Es importante comentar qué papel juega esta aplicación dentro del proyecto. Para dar una pequeña información, *treeTagger* es el programa que se va a encargar de analizar los “sentimientos” de cada *tweet*.

A medida que se han ido realizando diferentes búsquedas de información, se ha llegado a la conclusión, que hablar del modelo de *Markov*, de la aplicación *treeTagger* o *Part-of Speech* es hablar de un mismo concepto, todo está relacionado puesto que el punto de partida es *Markov*.

Por tanto antes de entrar dentro de *treeTagger* y *Part-of Speech* se va a explicar el principio básico, Modelo de *Markov*, del que parte todo lo relacionado con el análisis de textos y porqué y de qué manera surge el programa que se está utilizando, *treeTagger*.

2.3.1 Modelo de *Markov*

Antes de definir el modelo de *Markov*, se ha creído conveniente explicar ciertos conceptos que aparecen en la definición del mismo, para hacer que la definición del modelo sea un poco más entendible y así poder saber de qué se está hablando.

Modelo estadístico: Es la representación matemática, la cual se ha deducido de un conjunto de supuestos, que tienen el propósito de estudiar:

Los resultados que se obtienen en un experimento aleatorio y la predicción del comportamiento de este mismo experimento en un futuro.

Para que se obtengan resultados fiables es necesario, que ambos experimentos se realicen partiendo de las mismas condiciones iniciales.

Por otro lado, también permite conocer la distribución de probabilidades, que generarían los valores por los que pasa una variable. Asignándole como nombre Distribución de Probabilidad.

Proceso estocástico: proceso en el que se involucran probabilidades. Esto quiere decir que, en un proceso estocástico se van a representar todos los estados, que se necesitan para realizar una actividad, se representarán con el máximo detalle posible.

A cada paso generado se le asignara una probabilidad, este número indica que grado de probabilidad hay de que transite a ese estado.

Proceso de Markov: es un fenómeno aleatorio, que es dependiente del tiempo, para el cual se cumpla una propiedad específica. En un proceso



estocástico, sin memoria, la propiedad condicional sobre: presente, futuro y pasado del sistema son independientes. Los procesos de *Markov* surgen en probabilidad y estadística de una de las siguientes dos maneras:

a) La primera manera se basa en tomar un argumento por separado y mediante transformaciones matemáticas demostrar que tiene la propiedad de *Markov*, con esto se puede afirmar que tiene propiedad y se puede definir para todos los procesos de *Markov*.

b) Esta segunda manera es la más utilizada, en este caso se supone, que la propiedad es válida para un proceso aleatorio, con el fin de construir un modelo estocástico para ese proceso.

Mediante este método se definen muchas formas para introducir dependencia estadística a un modelo con un proceso estocástico. Gracias a esta característica se genera una dependencia entre los diferentes retardos, haciendo que disminuya a medida que el retardo aumenta.

Cadena de Markov: se usa para entender el proceso de *Markov*, se tiene que definir con estados finitos. Normalmente este tipo de cadenas están definidas para conjuntos discretos en el tiempo, aunque algunos autores lo han definido con valores continuos.

Por tanto, una vez explicados los conceptos ya se puede dar la definición formal del modelo de *Markov*.

El **modelo de Markov:** es un modelo estadístico en el que se supone que se va a analizar un proceso de *Markov*. En este caso las variables son ocultas, no son observables; para que el modelo siga adelante, son necesarias otras variables.

Estas nuevas variables son observables, por lo que se podrá obtener un resultado. El resultado serán los valores que se les asignarán a las variables ocultas.

En el momento en el que se tengan resultados para todas las variables (visibles u ocultas) se podrán realizar aplicaciones para el reconocimiento de patrones. Como se puede observar en la [Ilustración 9](#) una red bayesiana tiene mucha relación con un Modelo de *Markov*.

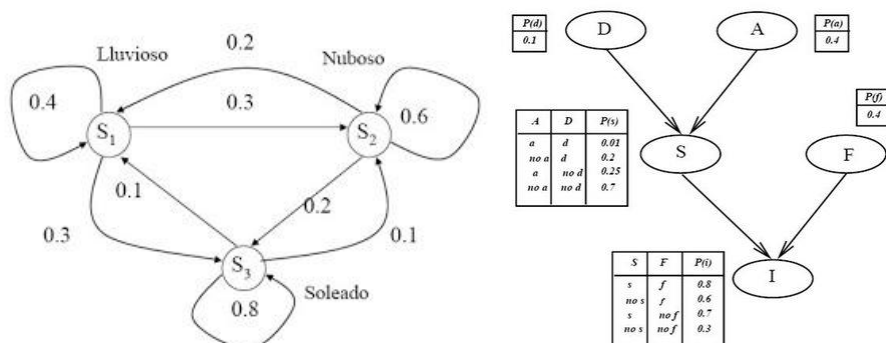


Ilustración 9. Red bayesiana dinámica simple y modelo de Markov

Como ya se ha comentado, el modelo de *Markov* se puede dividir en dos, a causa de los tipos que pueden tomar sus variables: ocultas o normales.

El modelo de Markov normal: En este caso todos los estados son visibles, por el observador, por lo que la incógnita está en averiguar la probabilidad que se le tiene que dar a cada estado para poder transitar entre ellos.

El modelo de Markov oculto: Para este modelo se van a tener los dos tipos de variables: ocultas y visibles. En este caso, si se tienen valores desconocidos, difícilmente se podrán sacar conclusiones, por lo que el objetivo es determinar qué valores puede tomar las variables ocultas, todo esto se tiene que hacer a partir de las variables visibles, puesto que es el único dato que se tiene. Por lo que se generará una distribución de probabilidad sobre los símbolos de la salida y así se obtendrá la información necesaria para poder obtener la secuencia de los estados.

Por la información recopilada, para el proyecto que aquí se presenta, el grupo más favorable es el segundo. Se ha llegado a esta conclusión puesto que se han encontrado numerosos programas que se encargan de diferentes tipos de reconocimientos como pueden ser: faciales, gestuales, del habla o etiquetado gramatical.

Por lo que con este último párrafo se puede afirmar que todo lo que se está comentado en este proyecto es válido y que el modelo de *Markov* lo sustenta.

2.3.2 ¿Qué es *TreeTagger*?

Es una herramienta creada por el instituto de Lingüística Computacional de la Universidad de Stuttgart, dentro del proyecto TC, el cual se encarga de anotar en textos con información, *part-of-speech* y lemas.



Gracias a la portabilidad de esta herramienta y su fácil adaptación, siempre que se tenga de un léxico y un corpus marcado, se ha podido utilizar en diferentes idiomas y textos.

Para conocer un poco más que es el *treeTagger* se han encontrado dos artículos, los cuales explican un poco mejor que es esta herramienta.

A continuación se va a realizar un pequeño resumen sobre cada uno de los dos artículos, exponiendo las ideas más importantes y que sean relevantes para este proyecto.

En el punto 9, se encuentra el link a cada uno de estos artículos. Los dos artículos han sido escritos por el mismo autor, *Helmut Schmid* profesor de la universidad de Stuttgart.

- El **primero artículo**, trata sobre la mejora que se ha realizado en la parte de *Part-of-Speech* para textos escritos en idioma alemán.

El motivo por el que se genera este artículo es porque a la hora de utilizar el *treeTagger* con otros idiomas, este no da los resultados óptimos, los cuales se esperan de él; esto ocurre puesto que todos los estudios realizados se han basado en el idioma inglés, por lo que a la hora de extrapolar a otros idiomas no todas las reglas definidas son válidas.

El primer problema, que se encontró, fue con la formación de las palabras, por lo que se ha ido completando el *treeTagger* con pequeños apéndices para que sea válido para cualquier idioma.

Como ya se ha comentado al principio de este punto el *treeTagger* está basado en el modelo de *Markov* por lo que el artículo comenta las similitudes con él y como se realiza la transformación estadística para que funcione con el análisis de sentimientos.

Para los árboles de decisión se plantearon dos formas para definir una palabra, comprobando la palabra anterior a esta o la posterior. Las dos opciones están basadas en el **algoritmo ID3**, se muestra el

árbol en la **Ilustración 10**, este algoritmo engloba la búsqueda de hipótesis o reglas dentro de él, mediante un grupo de ejemplos que se le proporcionan.

Los ejemplos tienen que estar compuestos por una serie de *tuplas* con valores, conteniendo a los atributos a clasificar que serán positivos o negativos.

Con estos ejemplos el algoritmo será capaz de clasificar y obtener hipótesis de nuevas instancias clasificándolas en positivas o negativas. Los resultados se verán en un árbol de decisión.

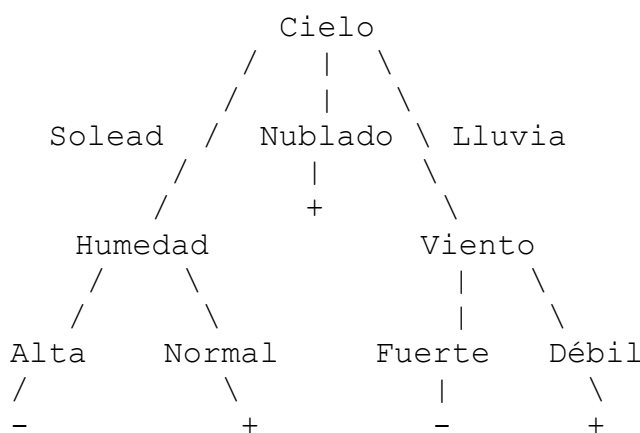


Ilustración 10. Árbol de decisión algoritmo ID3

Una vez explicado de qué forma se utiliza el algoritmo ID3, se pasa a comentar que módulos se han añadido para que el *treeTagger* se pueda utilizar para analizar cualquier texto en cualquier idioma.

Lo que intenta hacer es generar grupos de palabras que puedan empezar o terminar por lexemas parecidos y así poder realizar la agrupación de forma más global y no mediante un grupo fijo de palabras enteras, este es una de las cosas, las cuales se han cambiado puesto que no hay muchas palabras en inglés que empiecen por los mismos lexemas.

Para concluir, muestra una tabla en la que se representa el porcentaje de aciertos que se han obtenido al analizar diferentes textos, estos porcentajes se realizan a medida que se van creando nuevas versiones, en cada versión el autor introduce un nuevo cambio o mejora y mide el éxito.

Las conclusiones obtenidas de este artículo las muestra en la **Ilustración 11** y como dato a tener en cuenta se ha reducido casi un

37% los errores que cometía el *treeTagger* gracias a las nuevas implementaciones que se han realizado.

tagging method	accuracy
suffix lexicon only (1)	96.05 %
(1) + prefix lexicon	96.10 %
(1) + equival. class smoothing	96.52 %
(1) + sentence initial word treatm.	96.46 %
all features (5)	96.98 %
(5) + additional word/tag-pairs (6)	97.04 %
(6) + additional probabilities	< 97.04 %
(5) + standard MM formula	97.53 %

Ilustración 11. Resultados de las pruebas

- El **segundo artículo**, trata sobre la probabilidad de usar árboles de decisión en *Part-of-Speech* puesto que muchas palabras por si solas pueden tener varios significados, el ejemplo que indica es la palabra “store” en ingles puede ser un nombre, un infinitivo o un verbo, pero gracias al contexto de la frase se puede saber cuál de las tres opciones es la correcta.

Para poder realizar este análisis se han utilizado diferentes tipos de métodos probabilísticos, los cuales están basados en la primera o segunda orden del método de *Markov*, pero a causa de utilizar tantos parámetros, el método puede ser ineficientes, por lo que en este artículo se muestra otra opción para evitar este tipo de problemas y así crear un árbol de decisión sin ambigüedades.

Se crea un árbol de decisión con un máximo de tres niveles, se muestra en la [Ilustración 12](#), creándolo a través del algoritmo ID3 con ciertas modificaciones, ejecutándolo de forma recursiva. En cada vuelta se va comparando el resultado con el anterior y se va quedando con el mejor. (*Estas modificaciones son las que se nombran en el artículo anterior*).

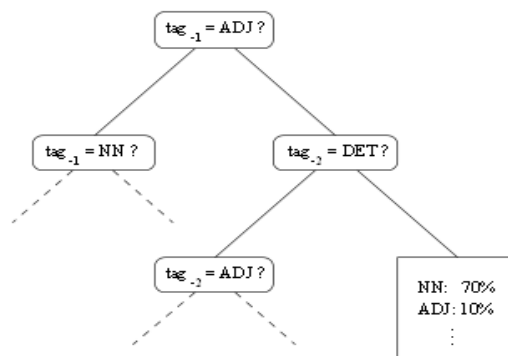


Ilustración 12. Recursividad del árbol

Para cada palabra le corresponden tres atributos: la palabra completa, sufijos y una entrada por defecto.

Si durante el proceso alguna palabra ya está definida se devuelven sus tres atributos. Para que no haya problemas, todas las palabras se pasan a minúsculas y se empieza con el análisis.

El análisis sigue utilizando el modelo de *Markov* para generar los árboles de decisión, pero se cambiaron ciertas ecuaciones de orden puesto que al hacer ese pequeño cambio suponían un aumento muy significativo en los aciertos.

En la **Ilustración 13** se muestra en la frecuencia de los nodos al realizar el análisis de “ess” como sufijo y dos nodos hijos “less” y “ness”. Es el resumen del árbol que se muestra en la **Ilustración 14**.

tag	suffix ess	suffix ness	suffix less
JJ	86	1	85
NN	10	2	8
NP	45	45	0
RB	2	0	2
total	143	48	95

Ilustración 13. Frecuencia

En la **Ilustración 14** se muestra un ejemplo con un árbol de tres niveles analizando diferentes sufijos.

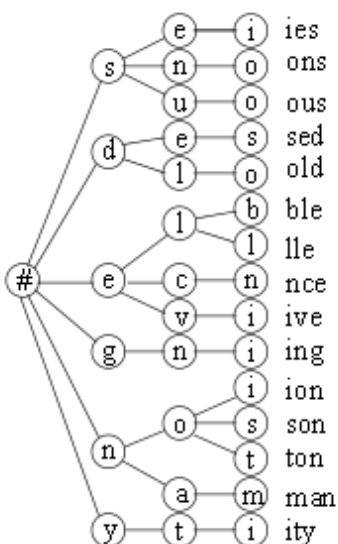


Ilustración 14. Árbol de 3 niveles

Como resultado del artículo se muestra, que realizando varios cambios se obtiene un acierto de 0.3% más sobre el estándar y un 0.6% sobre el

“*Bigram Version of TreeTagger*”. En la **Ilustración 15** se muestra resumido las probabilidades de éxito.

method	context	accuracy
trigram tagger	trigram	96.06 %
TreeTagger	bigram	95.78 %
TreeTagger (0.1)	trigram	96.34 %
TreeTagger	quatrogram	96.36 %
TreeTagger (10^{-10})	trigram	96.32 %

Ilustración 15. Probabilidades de éxito

El nuevo método es diferente al resto de métodos, la mayor diferencia se muestra a la hora de transmitir las probabilidades; una vez se han estimado las probabilidades, estas se mandan al árbol de decisión.

Gracias a las tablas queda demostrado que debido a los ajustes realizados, los resultados son mucho más satisfactorios.

Por otro lado se puede decir que *treeTagger* es robusto puesto que al introducir nuevas mejoras no baja su rendimiento y gracias a estas mejoras es capaz de analizar un número mayor de *tokens* por segundo.

2.3.3 Part of Speech

Para comprenderlo mejor el término *part-of-speech*, éste se refiere a una parte de la oración, en la que se incluye: palabras, clase o categoría léxica. Esto se define generalmente por el comportamiento sintáctico y morfológico que tenga el elemento léxico que se quiera comprobar.

Se realizan categorías lingüísticas comunes a sustantivos y verbos, con clases abiertas (se introducen palabras nuevas de forma regular) o clases cerradas (la frecuencia en la que se introducen nuevas palabras es baja o nula)

El problema empieza cuando entramos en los diferentes idiomas a pesar de que todos los idiomas tengan sustantivos, categorías verbales y adjetivos, este último difiere en gran medida según el idioma, en japonés tienen tres clases de adjetivos mientras que en inglés tienen solo uno.

Por otra parte los chinos, coreanos y japoneses tienen clasificadores nominales que las lenguas europeas no tienen y muchos otros idiomas ni siquiera tienen diferencia entre adjetivos, adverbios y verbos.



La clasificación de palabras en categorías léxicas se estableció entre los siglos V y VI AC y se basaba en cuatro categorías de palabras: nombres o sustantivos, verbos, prefijos y palabras invariables (preposiciones). Los cuales se agrupaban en dos grupos: flexivas (nombres y verbos) e inflexivas (prefijos y preposiciones).

A finales del siglo II AC este esquema se amplió a ocho categorías: Sustantivo, verbo, participio, interjección, pronombre, preposición, adverbio y conjunción.

Para poder aclarar un poco mejor las clases gramaticales, estas se van a disponer en la **Tabla 4** a modo de resumen.

Siglo V, VI		Siglo II	Hoy en día
Flexivas	Nombres	Sustantivos	Sustantivos
		Verbos	Verbos
		Participios	Participios
	Verbos	Interjecciones	Interjecciones
		Pronombres	Pronombres
Inflexivas	Prefijos	Preposiciones	Preposiciones
		Adverbios	Adverbios
	Preposiciones	Verbos	Adjetivos
			Verbos

Tabla 4. Resumen clases gramaticales

Inglés

Al idioma inglés cabe hacerle una mención especial puesto que es la base de la que parte tanto part of speech como treeTagger.

En la **Ilustración 16** se muestra un diagrama en forma de árbol con las categorías que componen el idioma en los estudios lingüísticos de hoy en día.

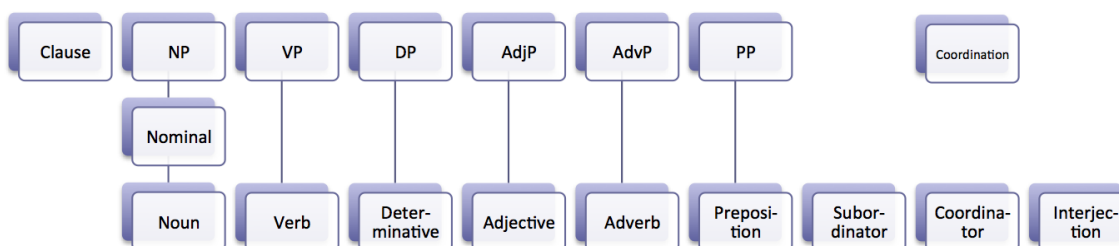


Ilustración 16. Diagrama categorías lingüísticas inglesas

A pesar de tener esas ocho categorías tradicionales, los lingüistas modernos han sido capaces de agruparlas en cuatro partes principales como son: sustantivos, verbos, adjetivos y adverbios. Este formulario que han creado es por las similitudes de la palabra por su pronunciación y escritura en cada parte de la oración.



Parece sencillo el hecho de definir una palabra, al menos en español, pero no en todos los idiomas se tiene la misma facilidad, puesto que una misma palabra puede tomar diferentes valores según en el contexto en el que se utilice, (nombre, adjetivo, adverbio, verbo) por lo que es necesario primero saber qué significado tiene dentro de la frase y luego posicionarla, en este caso, se cree que este tipo de palabras no se tendrían que posicionar en ningún lugar, puesto que una vez que se posicionen, puede ser que siempre se asignen al mismo grupo y eso no sería correcto.

2.3.4 *TreeTagger vs. Connexor*

Buscando información sobre *treeTagger* se ha encontrado un artículo que lo compara con otro analizador de textos, *connexor*. Según mi trato con *treeTagger* y lo que se ha leído en el artículo, se ha llegado a las siguientes conclusiones:

- **TreeTagger** fue la primera herramienta que se utilizó para análisis de textos. Como se ha podido observar durante este proyecto, la herramienta cumple en gran medida su función y es muy competente. A todo esto hay que sumar que es una herramienta de libre uso, no hay que pagar por utilizarla. Otra ventaja de *treeTagger* es que es más portable a otros idiomas que *connexor*.
- **Connexor** por otra parte fue creada a partir de *treeTagger* y es privada, por lo que hay que pagar para poder disfrutar de su análisis. En cualquier caso la diferencia no solo es el precio, sino que también llega a analizar las relaciones sintácticas que existen entre los dos *tokens*. También cabe destacar que el resultado que obtienes mediante *connexor* es algo más elaborado que con *treeTagger* puesto que analiza los textos en función de:
 - La posición de la palabra en la oración
 - La palabra
 - El tipo de raíz
 - La relación sintáctica
 - Las etiquetas sintácticas y morfológicas.

A pesar de dar más información sobre la salida del texto o de qué forma se ha analizado, para este proyecto creo que la información que da *treeTagger* es mucho más útil y más clara que la que da *Connexor*.

Además los fallos que se pudieran tener en el análisis de textos de otros idiomas *treeTagger* ha sido capaz de superarlos, mientras que *Connexor* los sigue teniendo.

Para concluir, para ser un producto de pago, creo que *Connexor*, no da los resultados esperados teniendo sobre todo si se tiene en cuenta que hay otra herramienta libre que sí que los proporciona.

2.4 Google y Bing



Puede sorprender encontrar dos buscadores en este documento, puesto que no se tiene porqué entender muy bien qué función tienen dentro del análisis de sentimientos.

Como se ha podido ver se ha seguido una estructura parecida en cada concepto dentro del punto 2

ESTADO DEL ARTE, explicando que era cada concepto y luego la función que desempeñan dentro del proyecto; pero en este caso no se cree que sea necesario explicar que son Google y Bing pero sí que funciones de ellos se utilizan y así comprender la finalidad que tienen dentro de este proyecto, la cual es bastante importante.

2.4.1 Finalidad

En este punto se va a explicar una parte bastante importante del proyecto puesto que se introduce una nueva tecnología, en este caso estos dos buscadores ayudarán con el análisis de sentimientos, su cometido será el de poder discernir si un adjetivo es positivo o negativo.



Para el ser humano muchas veces parece una obiedad el saber si un adjetivo es positivo o negativo, pero aún así en muchas ocasiones según en el contexto en el que se utilice el adjetivo puede llegar a significar cosas diferentes.

Por tanto si ya para nosotros puede ser una cosa un tanto ambigua para decidir, imaginemos una máquina, si no le damos las pautas necesarias nunca será capaz de diferenciarlo, por ese motivo entran en juego los dos buscadores.

En el punto 4.1 se comentará con que finalidad se van a utilizar estos adjetivos. En este punto nos interesa comentar y explicar como se realiza la comparacion entre adjetivos.





CAPÍTULO 3

Análisis

En este capítulo se describirá la fase de análisis del proyecto, incluyendo los requisitos software y hardware de la aplicación y los distintos casos de uso.





3 ANÁLISIS

3.1 Visión general del sistema

Antes de comenzar a exponer los requisitos, que tiene que cumplir la aplicación, se intentará explicar cuál es la finalidad del proyecto.

En este proyecto se define una página web en la cual se podrán buscar *tweets* con unos requisitos concretos, los cuales se clasificarán según los adjetivos y estructuras sintácticas que compongan cada *tweet*, una vez realiza esta clasificación se mostrarán los resultados de diferentes formas.

Esta clasificación se denominará sentimiento; este sentimiento es en primer lugar decidir si el adjetivo por si solo es positivo o negativo y en segundo lugar es decidir con el conjunto de todos los adjetivos, si el *tweet* tiene un sentimiento positivo, negativo o neutro.

Una vez se tengan todos los *tweets* analizados, se pasará a mostrar los resultados de diferentes maneras:

- **Zona de texto con *scrollbar***, en esta parte se mostrarán uno a uno todos los *tweets* que se han analizado, junto al *tweet* aparecerá una imagen, que representará de una forma gráfica que tipo de sentimiento se le ha asignado, pudiendo ser: positivo con el icono será verde, negativo (rojo) o neutro (amarillo).
- **Zona resumen**, en esta parte se presentarán los tres tipos de sentimientos y a continuación se indicará el número de *tweets* que hay de cada sentimiento.
- **Zona resumen gráfico**, este resumen indicará mediante una barra de progresión que sentimiento tiene la marca de forma global.

3.2 Requisitos Funcionales

Los requisitos se han dividido en requisitos funcionales y no funcionales. Estas dos categorías especifican qué características debe tener la aplicación (requisitos funcionales) y cómo se deben hacer las cosas en cada parte del sistema (requisitos no funcionales) para que funcione todo tal y como se ha especificado.

A continuación se va a explicar el formato de la tabla y la explicación de cada campo.



Identificador	TipoRequisito- XXX
Nombre	
Descripción	
Necesidad	
Dependencia	

Tabla 5. Formato de la tabla

- **Identificador:** es un elemento identificativo único, siempre va a tener el mismo formato TipoRequisito-XXX. Dentro del TipoRequisito se pueden dar los siguientes valores: RF, RNFIIE, RNFIU, RNR y las XXX equivale a 3 números consecutivos empezando en 001 siempre que el TipoRequisito se modifique.
- **Nombre:** nombre del requisito. Deberá resumir la finalidad del requisito y debe empezar siempre que se pueda con un verbo en infinitivo.
- **Descripción:** explicación que se da sobre el requisito, permitiendo dar más información sobre la finalidad/funcionalidad de ese requisito para que el programador no tenga que realizar ninguna interpretación.
- **Necesidad:** cuanto se necesita ese requisito para la implementación del proyecto. Las opciones que puede tener este campo son: alta, media o baja.
- **Dependencias:** Se indicaran el/los identificador/es de los requisitos que sean depende para poder ser ese requisito implementado. Si se encuentra N/A quiere decir no aplica.

Identificador	RF- 001
Nombre	Recibir marca
Descripción	El usuario tiene que introducir la marca en el <i>input</i> que se presenta a la izquierda del botón.
Necesidad	Alta
Dependencia	N/A

Tabla 6. RF 001 – Recibir marca

Identificador	RF- 002
Nombre	Buscar tweets en función de la marca
Descripción	A través del Api de Twitter se conseguirán todo los tweets de la marca introducida.
Necesidad	Alta
Dependencia	RF-001

Tabla 7. RF 002 – Buscar tweets en función de la marca



Análisis de sentimientos acerca de marcas mediante Tweets

Identificador	RF- 003
Nombre	Identificar adjetivos
Descripción	Una vez que se tiene un tweet se utiliza el api de treeTagger para poder catalogar las palabras y así conseguir los adjetivos.
Necesidad	Alta
Dependencia	RF-002

Tabla 8. RF 003 – Identificar adjetivos

Identificador	RF- 004
Nombre	Valorar tweets
Descripción	Dividir y contabilizar el número de adjetivos positivos y negativos para darle un sentimiento al tweet.
Necesidad	Alta
Dependencia	RF-003 y RF-005

Tabla 9. RF 004 – Valorar tweets

Identificador	RF- 005
Nombre	Clasificar adjetivos
Descripción	Comprobar si el adjetivo a valorar se tiene definido o no. Si está definido analizar directamente si no lo está utilizar api de Google. Una vez que se tenga definido pasar a clasificar en positivo o negativo
Necesidad	Alta
Dependencia	RF-003

Tabla 10. RF 005 – Clasificar adjetivos

Identificador	RF- 006
Nombre	Valorar marca
Descripción	Sumar y multiplicar por un peso el número de tweets clasificados en: positivo negativo y neutro, sumar los valores y dar el resultado.
Necesidad	Alta
Dependencia	RF-004

Tabla 11. RF 006 – Valorar marca

Identificador	RF- 007
Nombre	Mostrar resultado
Descripción	Diferentes formas de mostrar los resultados: <ul style="list-style-type: none">▪ Tweet: uno a uno junto con su sentimiento▪ Global: valoración de la marca▪ Por Sentimineto: agrupar los tweets analizados.
Necesidad	Alta
Dependencia	RF-006

Tabla 12. RF 007 – Mostrar resultado



3.3 Requisitos No Funcionales

3.3.1 Interfaces externas

Los requisitos que se muestran en este apartado van relacionados con elementos externos que se necesitan para que la aplicación funcione.

En este caso se tiene que utilizar el Api de *Twitter*, para conseguir los *tweets*, el Api de *TreeTagger* para catalogar las palabras y el Api de *Google* para buscar los sentimientos de los adjetivos que no se saben cómo analizar.

Identificador	RNFIE- 001
Nombre	Api de Twitter
Descripción	Se necesita para saber que método utilizar para sacar los tweets escritos en español y con el nombre de la marca
Necesidad	Alta
Dependencia	N/A

Tabla 13. RNFIE 001 – Buscador de Tweets

Identificador	RNFIE- 002
Nombre	Api de Google
Descripción	Se necesita para obtener los resultados al realizar la función AROUND y ver si es válida para el proyecto.
Necesidad	Alta
Dependencia	N/A

Tabla 14. RNFIE 002 –Api de Google

Identificador	RNFIE- 003
Nombre	Librería de treeTagger
Descripción	Se necesita para obtener las categorías gramaticales de las palabras que componen el tweet.
Necesidad	Alta
Dependencia	N/A

Tabla 15. RNFIE 003 – Librería de treeTagger

3.3.2 Interfaz de usuario

Identificador	RNFIU- 001
Nombre	Todos los sistemas operativos
Descripción	La aplicación va a ser portable y será válida para cualquier sistema operativo en el que se ejecute.
Necesidad	Alta
Dependencia	N/A

Tabla 16. RNFIU 001 – Todos los sistemas operativos



Análisis de sentimientos acerca de marcas mediante Tweets

Identificador	RNFIU- 002
Nombre	TreeTagger
Descripción	Es necesario tener instalado el ejecutable de treeTagger en el equipo
Necesidad	Alta
Dependencia	N/A

Tabla 17. RNFIU 002 – TreeTagger

Identificador	RNFIU- 003
Nombre	Accesibilidad para personas mayores
Descripción	Se han utilizado letras grandes y colores básicos para que las personas mayores puedan utilizarlo sin problema.
Necesidad	Alta
Dependencia	N/A

Tabla 18. RNFIU 003 – Accesibilidad para personas mayores

Identificador	RNFIE- 004
Nombre	Accesibilidad para usuarios principiantes
Descripción	La aplicación es intuitiva y no se necesita de conocimientos previos para su utilización.
Necesidad	Alta
Dependencia	N/A

Tabla 19. RNFIU 004 – Accesibilidad para usuarios principiantes

Identificador	RNFIE- 006
Nombre	Conexión a internet
Descripción	Es necesario que el equipo esté conectado a internet. Para poder conseguir los tweets y las respuestas de google.
Necesidad	Alta
Dependencia	N/A

Tabla 20. RNFIU 006 – Conexión a internet

Identificador	RNFIE- 007
Nombre	Portátiles y PC
Descripción	Es válido en cualquier tipo de PC o portátil.
Necesidad	Alta
Dependencia	N/A

Tabla 21. RNFIU 007 – Portátiles y PC

3.3.3 Requisitos de rendimiento

Identificador	RNR- 001
Nombre	Número máximo de tweets
Descripción	Se analizarán como máximo los 100 tweets más recientes.
Necesidad	Media
Dependencia	N/A

Tabla 22. RNR 001 – Número máximo de tweets



Identificador	RNR- 002
Nombre	Tiempo máximo de respuesta
Descripción	El tiempo medio de espera tiene que ser de 5 minutos, va a influir las veces que se tenga que ir al API de google para definir el adjetivo.
Necesidad	Media
Dependencia	N/A

Tabla 23. RNR 001 – Número máximo de tweets

3.4 Casos de usuario

En este apartado se va a comentar los casos de uso de dos formas: con una descripción gráfica, en la que se mostrarán como interactúa el usuario con la aplicación y una descripción textual en la que se explica en profundidad cada una de las opciones que puede realizar el usuario.

3.4.1 Descripción gráfica

En la **Ilustración 17** se observa al usuario y las dos opciones, que puede realizar dentro de la aplicación: Activar marca y ver resultados.

En la opción de **Activar Marca**, se introduce el nombre de la marca a analizar y se presiona el botón de buscar.

En la opción de **Ver Resultados**, en esta opción se incluyen las tres formas, en las que se van a mostrar los resultados, una vez hayan sido analizados los *tweets* y la marca.

En **Tweet** se mostrarán cada uno de los *tweets* analizados junto con su sentimiento.

En Global se mostrará la valoración de la marca.

Por Sentimiento se mostrarán el número de *tweets* que hay de cada sentimiento.

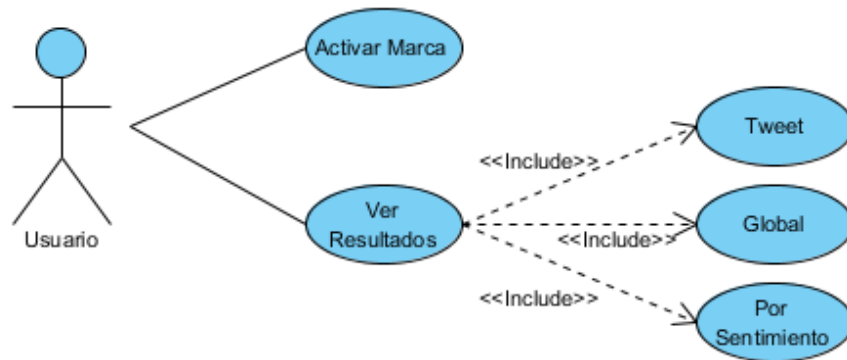


Ilustración 17. Casos de uso

3.4.2 Descripción Textual

Para poder explicar de forma textual el diagrama de casos de uso, se ha creado la **Tabla 24** para poder explicar detalladamente cada caso de uso. Primero se mostrará el esquema de la tabla y seguidamente la explicación de cada campo:

Identificador	CU-YYY
Nombre	
Actores	
Objetivo	
Descripción	
Pre-condición	
Post-condición	

Tabla 24. Formato de la tabla de requisitos

- **Identificador:** identifica a cada caso de uso. El formato utilizado es el siguiente “**CU YYY**”, donde cada “**Y**” se corresponde con un número comprendido entre 0 y 9.
- **Nombre:** se corresponde con el nombre asignado al caso de uso.
- **Actores:** son los distintos roles de usuarios que encontraremos en el sistema. Es cualquier entidad externa al Sistema que está relacionada con éste y que le demanda una funcionalidad.
- **Objetivo:** se corresponde con la finalidad que se pretende obtener tras realizarse el caso de uso.
- **Descripción:** se corresponde con la descripción del caso de uso referido.



Análisis de sentimientos acerca de marcas mediante Tweets

- **Pre-condición:** se corresponde con las condiciones iniciales que se deben cumplir para que se realice dicho caso de uso.
- **Post-condición:** se corresponde con las condiciones finales que se cumplirán tras realizar dicho caso de uso.

A continuación se muestran todos los casos de uso desarrollados:

Identificador	CU-001
Nombre	Pedir valoración
Actores	Usuario
Objetivo	Iniciar en el sistema la valoración de la marca introducida.
Descripción	<ul style="list-style-type: none">▪ Usuario tiene que introducir el nombre de la marca en el <i>input</i>▪ Pulsar el botón “Buscar”
Pre-condición	El sistema tiene que estar listo para recibir la marca a valorar
Post-condición	La valoración de la marca

Tabla 25. CU 001 – Pedir valoración

Identificador	CU-002
Nombre	Ver Resultados
Actores	Usuario
Objetivo	Mostrar de diferentes formas el resultado de la valoración de la marca
Descripción	<ul style="list-style-type: none">▪ El sistema mostrará en diferentes puntos de la interfaz diferentes tipos de resultados:<ul style="list-style-type: none">▪ <i>Tweets</i>: cada <i>tweet</i> con su sentimiento▪ Por Sentimientos: agrupados por sentimientos▪ Global: valoración de la marca
Pre-condición	El sistema tiene que estar listo para recibir los resultados
Post-condición	Valoración de la marca

Tabla 26. CU 002 – Ver Resultados

Identificador	CU-003
Nombre	<i>Tweet</i>
Actores	Usuario
Objetivo	Mostrar cada <i>tweet</i> con su sentimiento.
Descripción	<ul style="list-style-type: none">▪ El usuario puede navegar por el panel de texto con scrollbar para ver todos los <i>tweets</i> y sus sentimientos
Pre-condición	El sistema tiene que estar listo para recibir la lista de <i>tweets</i> con sus sentimientos
Post-condición	La valoración de cada <i>tweet</i>

Tabla 27. CU 003 – *Tweet*



Análisis de sentimientos acerca de marcas mediante Tweets

Identificador	CU-005
Nombre	Global
Actores	Usuario
Objetivo	Mostrar la valoración de la marca
Descripción	<ul style="list-style-type: none">▪ En la barra de progresión se mostrará la valoración global de la marca. Esta barra se dividirá en tres partes.<ul style="list-style-type: none">▪ Primer tercio: Malos sentimientos hacia la marca▪ Segundo tercio: La marca no es atractiva▪ Tercer tercio: Buenos sentimientos hacia la marca
Pre-condición	El sistema tiene que estar listo para recibir la marca a valorar
Post-condición	La valoración de la marca

Tabla 28. CU 005 – Global

Identificador	CU-004
Nombre	Por sentimiento
Actores	Usuario
Objetivo	Iniciar en el sistema la valoración de la marca introducida.
Descripción	<ul style="list-style-type: none">▪ Mostar el nombre de cada tipo de sentimiento junto con el número de <i>tweet</i> de ese sentimiento y una barra que lo indicará de forma gráfica.
Pre-condición	El sistema tiene que estar listo para recibir los <i>tweets</i> por sentimiento
Post-condición	La valoración por sentimiento

Tabla 29. CU 004 – Por sentimiento





CAPÍTULO 4

Diseño

En este capítulo se destallará la solución de diseño escogida mediante la cual satisfacer todos los requisitos y casos de uso recogidos en la fase de análisis. También se abordará el diseño de la interfaz que se ofrecerá al usuario.





4 DISEÑO

En este capítulo se va a explicar de qué forma se ha realizado el prototipo de la aplicación y cómo funciona la aplicación, junto con las tecnologías utilizadas.

La aplicación se encarga de dar la valoración de una marca, en función de los comentarios, que realizan los usuarios a través de *Twitter*

Para analizar estos *tweets* se necesita la ayuda de diversas API's de programas externos a este proyecto.

1. API de *Twitter*: para poder conseguir los *tweets* con el nombre de la marca.
2. API de *TreeTagger* : para poder clasificar las palabras, que integran el *tweet* y así coger las que interesan al programa, en este caso los adjetivos.
3. API de *Google*: para valorar los adjetivos que no se hayan tenido en cuenta previamente.

Con la unión de los servicios que proveen estas API's se obtiene la valoración de la marca. A continuación se mostrará de qué forma se va disponer la información relativa a la valoración:

1. Por *tweet*, se mostrarán todos los *tweets* junto con su sentimiento.
2. Por sentimiento, se mostrarán el número de *tweets* que hay en cada tipo de sentimiento.
3. Global, se mostrará la valoración global que se le da a la marca, en función de la valoración que se haya obtenido gracias al análisis que se ha realizado de los *tweets*.

4.1 Funcionamiento del sistema

En la **Ilustración 18** se muestra de forma gráfica los distintos pasos que se seguirán para, a partir de una determinada marca introducida, conseguir obtener su valoración global según los *tweets* en los que se menciona. Tal y como se muestra en la imagen el proceso asociado al análisis de sentimientos se desarrollará mediante 5 fases:



Análisis de sentimientos acerca de marcas mediante Tweets

- 1.- Fase 1 – Inicio: El usuario introduce un término de búsqueda correspondiente a una marca y dispara la búsqueda
- 2.- Fase 2 – Recuperación: Se recuperan los *tweets* en los que se menciona dicha marca
- 3.- Fase 3 – Identificación de adjetivos: se identifican los adjetivos que aparecen en el *tweet*
- 4.- Fase 4 – Procesado de adjetivos: Se comparan los adjetivos encontrados en el *tweet* con dos listas de adjetivos: adjetivos con connotación positiva y adjetivos con connotación negativa. Si el adjetivo no aparece en ninguna de las dos listas se dispara un proceso auxiliar de análisis de sentimiento para asignarlo a una de ellas.
- 5.- Fase 5 – Sentimiento de la Marca: Valoración global de toda la marca gracias a la valoración de cada *tweet*.

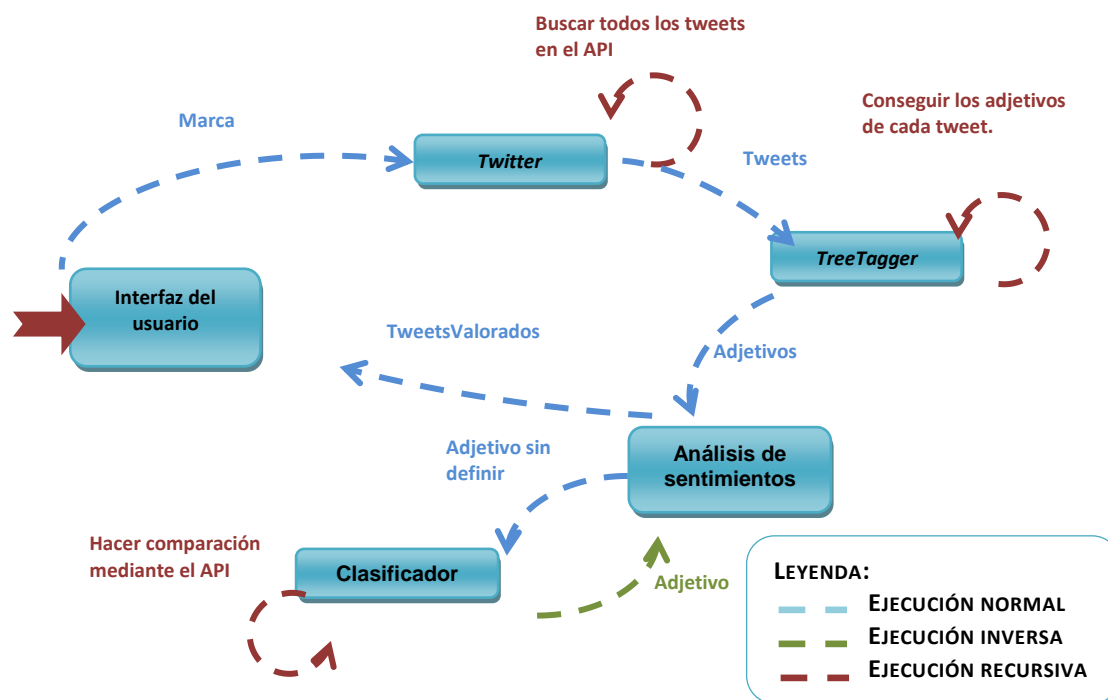


Ilustración 18. Funcionamiento del sistema

A continuación se detallará cada una de las fases.

4.1.1 Fase 1 – Inicio

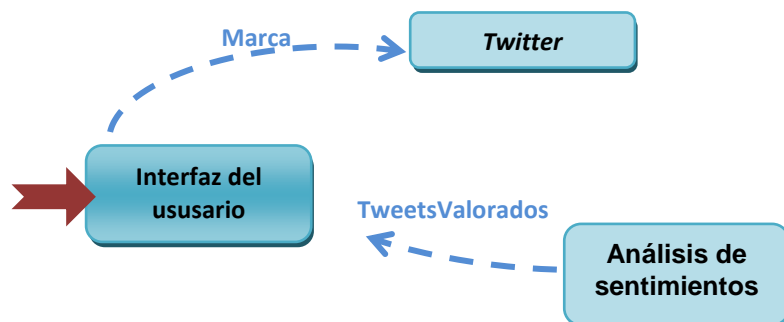


Ilustración 19. Funcionamiento del sistema - Fase 1

La primera fase comprende la de captura de información del usuario y el lanzamiento del procesamiento de *tweets* para obtener la aplicación. Para ello el interfaz de la aplicación deberá incluir campos de texto y botones que permitan al usuario introducir el nombre de la marca, activar la búsqueda.

4.1.2 Fase 2 – Recuperación Tweets

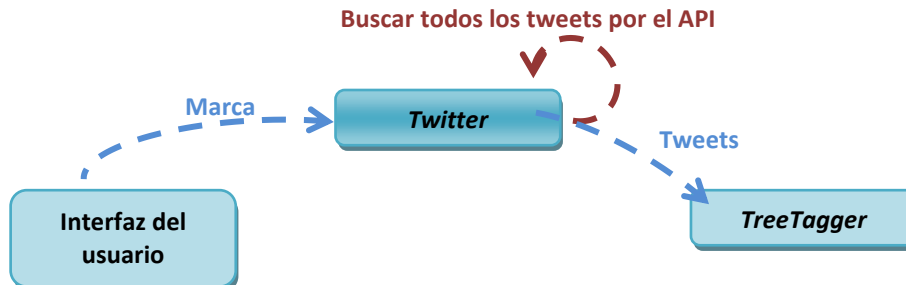


Ilustración 20. Funcionamiento del sistema - Fase 1

Esta fase es oculta para el usuario, se disparará al pulsar el botón si la marca se ha introducido previamente.

Para obtener los *tweets* es necesaria la ayuda del api de *Twitter*. Lo importante a la hora de realizar la consulta es la inicialización de las siguientes variables:

q= es el que define que nombre de marca que se va a buscar

lang= es el idioma en el que están escritos los *tweets*.

4.1.3 Fase 3 – Identificación de adjetivos

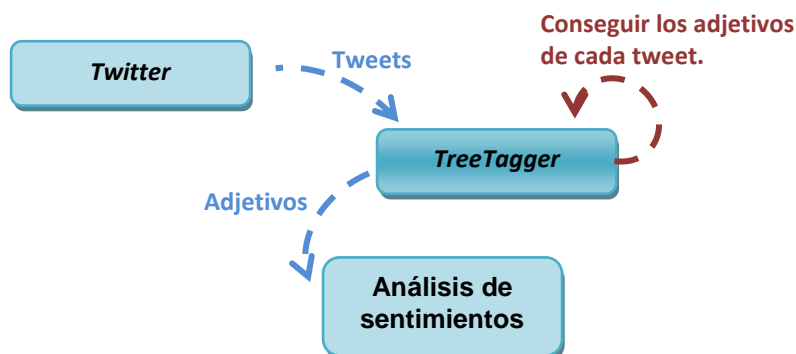


Ilustración 21. Funcionamiento del sistema - Fase 3

Esta fase será oculta para el usuario; una vez que se recibe el fichero en el que se encuentran todos los tweets guardados, se irá recorriendo el fichero para ir recogiendo la información contenida en él. Por lo que se irá cogiendo *tweet* a *tweet*, aplicando el siguiente procedimiento a cada uno de ellos:

1. Se envía el *tweet* a la función *treeTragger*, la cual se de coger cada palabra del *tweet* y asociarla a su categoría gramatical.
2. Elección de adjetivos. Una vez procesadas todas las palabras del *tweet*, se seleccionarán únicamente aquellas a las que se les haya asignado ADJ (adjetivo)
3. Una vez obtenidos todos los adjetivos del *tweet* se pasará a analizarlos uno por uno. Durante el proceso de análisis pueden ocurrir varias cosas:
 - Que se sepa cómo analizarlo, porqué se encuentre dentro de una estructura o sea un adjetivo básico ya definido, esta parte se explicará en el punto [4.1.4](#).
 - Que no se sepa cómo analizarlo, por lo que se pasará a utilizar la opción de los buscadores, esta parte se explicará en el punto [4.2](#).

4.1.4 Fase 4 – Análisis de adjetivos

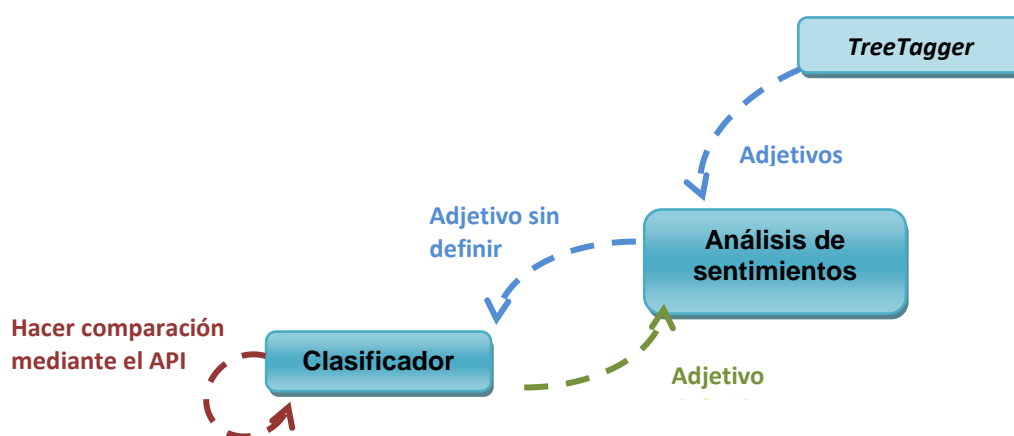


Ilustración 22. Funcionamiento del sistema - Fase 4

El análisis de adjetivos se ha dividido en dos pasos: análisis del adjetivo y resultado del *tweet*.

El paso de **análisis del adjetivo** se realizará por cada adjetivo encontrado en el *tweet*, tiene por objeto determinar si el adjetivo forma parte de una estructura, es comparativo, superlativo o es un adjetivo básico:

Adjetivos que forman parte de una estructura

- En la escritura en español, en muchas ocasiones se utilizan los adjetivos dentro de estructuras para dar mayor connotación a la frase y así poder cargar de significado del adjetivo. En la [Tabla 30](#) se han recogido las estructuras que se han tenido en cuenta para el análisis.

Sentimiento	Expresión
Positivo	Más _____ que
Negativo	Menos _____ que
Neutro (igual)	Tan _____ como

Tabla 30. Estructuras básicas

Adjetivos que forman parte de una comparación

- Con estos adjetivos ya se encuentra una fuerte connotación, por lo que el sentimiento, en cierto modo, ya va implícito en el adjetivo elegido. Los adjetivos comparativos de desigualdad que se han elegido para el análisis se muestran en la [Tabla 31](#).



Base	Comparativo
De bueno	Mejor
De malo	Peor
De grande	Mayor
De pequeño	Menor

Tabla 31. Adjetivos comparativos

Adjetivos superlativos

- En este caso se tiene que hacer dos distinciones, si el adjetivo tiene sufijo o si va acompañado de un adverbio. En la [Tabla 32](#) se muestra un ejemplo de las terminaciones de los adjetivos que se han tenido en cuenta.

Expresión	Ejemplo
-ísimo	Buenísimo, malísimo o pésimo
-blísimo	Agradabilísimo
-able	Agradable
Adverbios + adjetivos	Muy bueno, muy sucio o sumamente interesante.

Tabla 32. Superlativos absolutos

Adjetivo básico

- En este caso se contará con dos grupos de adjetivos: uno de adjetivos con connotaciones positivas y otro de adjetivos con connotaciones negativas.

Inicialmente estos dos grupos incluirán los adjetivos básicos más comunes y utilizados a la hora de escribir, aunque como es lógico, no todos los adjetivos estarán presentes en estas dos listas.

Por ello, a medida que se vayan encontrando nuevos adjetivos que no aparezcan en ninguna de las dos listas se calculará su connotación siguiendo los pasos del algoritmo descrito en el punto [4.2](#) y se añadirán al correspondiente grupo.

De esta forma el tamaño de los grupos irá aumentando progresivamente cuanto más se utilice la aplicación y más búsquedas se lleven a cabo

Una vez que se ha terminado con el **análisis del adjetivo** se tiene que guardar el resultado obtenido, para así poder continuar con los siguientes adjetivos del *tweet*, si los hubiera.

Para ello se utilizarán dos contadores (positivo y negativo) que acumularán el número de veces que aparecen adjetivos con connotaciones positivas y negativas respectivamente dentro de un *tweet*

A la finalización del análisis se realizará una comparación entre el valor de los contadores. Si el resultado es favorable, mayoría de positivos, se considerará el *tweet* como positivo, si es desfavorable, mayoría de negativos, el *tweet* será negativo y si la resta de los dos contadores es cero el resultado es neutro.

Este nuevo resultado se tiene que guardar para poder utilizarlo en la valoración global de la marca. Los *tweets* se van almacenando hasta que se han analizado todos y finalmente se mostrarán junto con su sentimiento asignado.

4.1.5 Fase 5 – Sentimiento de la Marca



Ilustración 23. Funcionamiento del sistema - Fase 5

Una vez se han procesados todos los *tweets*, se lleva a cabo la valoración global de la marca. El resultado final se presentará de dos formas distintas:

- Resultado desglosado por sentimiento: se visualizará el valor de los 3 contadores: número de *tweets* favorables, número de *tweets* desfavorables y número de *tweets* neutros.
- Resultado global de la marca, que se calculará como el resultado de multiplicar cada contador por un peso (positivo 2, negativo: -1 y neutro: 1) y sumar los valores resultantes.

4.2 Algoritmo de Obtención de Connotación del Adjetivo

Este proceso se utiliza para obtener el tipo de connotación de un adjetivo que es encontrado por el sistema por primera vez y que por tanto no está incluido en ninguna de las listas base.

Para obtener dicha connotación se lanzarán búsquedas en *Google* que incluirán el adjetivo a definir, adjetivos de los dos grupos cuyas connotaciones han sido establecidos y operadores de búsqueda *Around* y *Near*.

Las búsquedas se realizarán únicamente sobre los 5 primeros adjetivos de cada grupo y de forma alterna.

Es decir, que si se empieza con un adjetivo de la lista de positivos, el siguiente adjetivo a utilizar sería su correspondiente en la lista de negativos. El resultado de estas búsquedas será el número de páginas en las que ambos adjetivos han sido encontrados cerca.

Dichos números se compararán, de tal forma que siempre se seleccionará el más alto y, dependiendo de si se corresponde con el adjetivo positivo o el negativo, se aumentará un contador de resultados positivos o negativos, respectivamente.

En el momento en el que se acaben ambas listas de adjetivos (buenos y malos) se restarán los contadores, y según sea el resultado obtenido se añadirá el resultado al grupo de adjetivos positivos o negativos. La **Ilustración 24** muestra de forma esquemática el proceso.

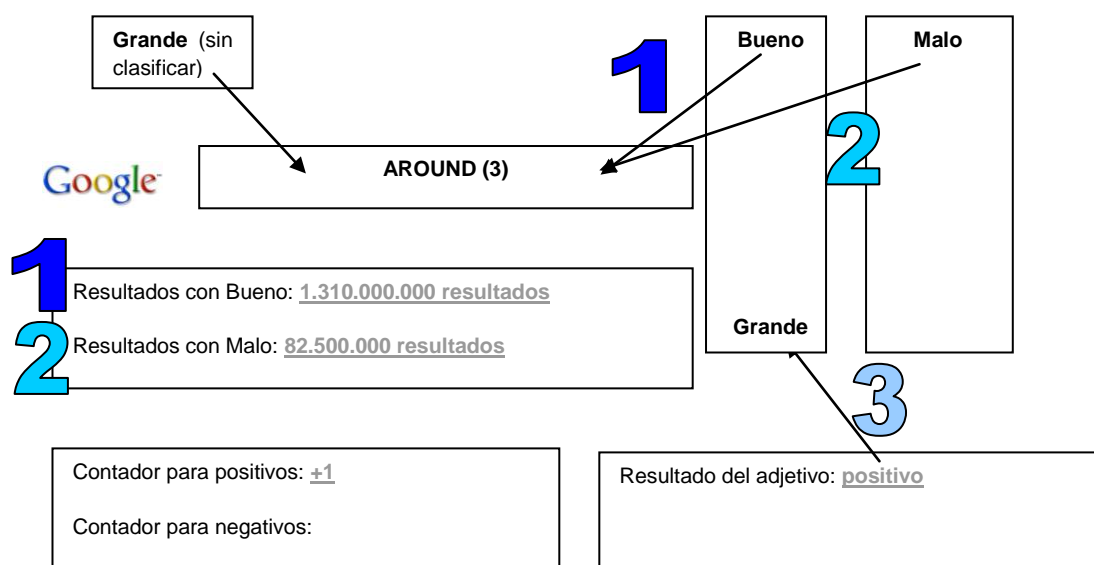


Ilustración 24. Explicación de los buscadores Google y Bing

Para probar que este algoritmo funciona, se van a mostrar varias pantallas realizando las búsquedas en el propio *Google*, para así mostrar que resultados se obtienen e intentar explicarlo mejor.

Se va a realizar la comparación con el adjetivo *sucio* para ver qué resultado se obtiene. Se va a ejecutar el algoritmo con 4 pares de adjetivos: guapo – feo, calor – frío, bien – mal y lleno – vacío.



Comparamos *sucio* con el par guapo – feo



Ilustración 25. Resultado de la comparación sucio con guapo – feo

Resultado para sucio – guapo = 5.730.000

Resultado para sucio – feo = 9.200.000

Contador de positivos: 0 Contador de negativos: 1

Resultado provisional: negativo

Comparamos *sucio* con el par calor – frío

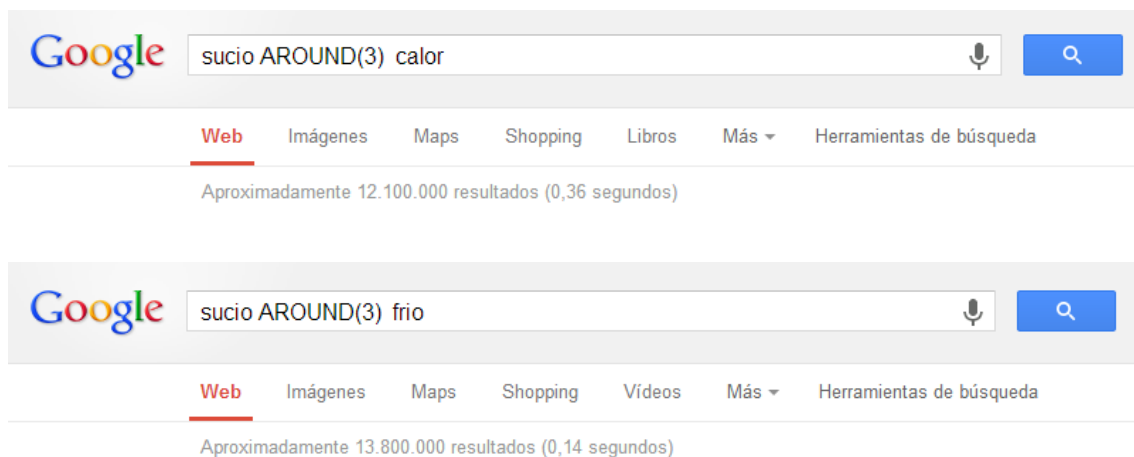


Ilustración 26. Resultado de la comparación sucio con calor – frío

Resultado para sucio – guapo = 12.100.000

Resultado para sucio – feo = 13.800.000

Contador de positivos: 0 Contador de negativos: 2

Resultado provisional: negativo



Comparamos *sucio* con el par bien – mal

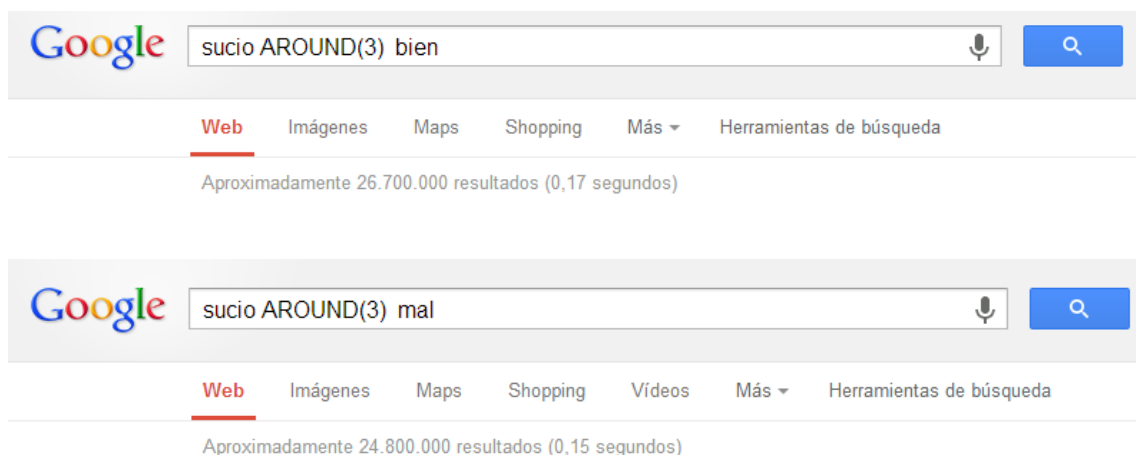


Ilustración 27. Resultado de la comparación sucio con bien – mal

Resultado para sucio – bien = 26.700.000

Resultado para sucio – mal = 24.800.000

Contador de positivos: 1 Contador de negativos: 2

Resultado provisional: negativo

Comparamos *sucio* con el par lleno – vacío

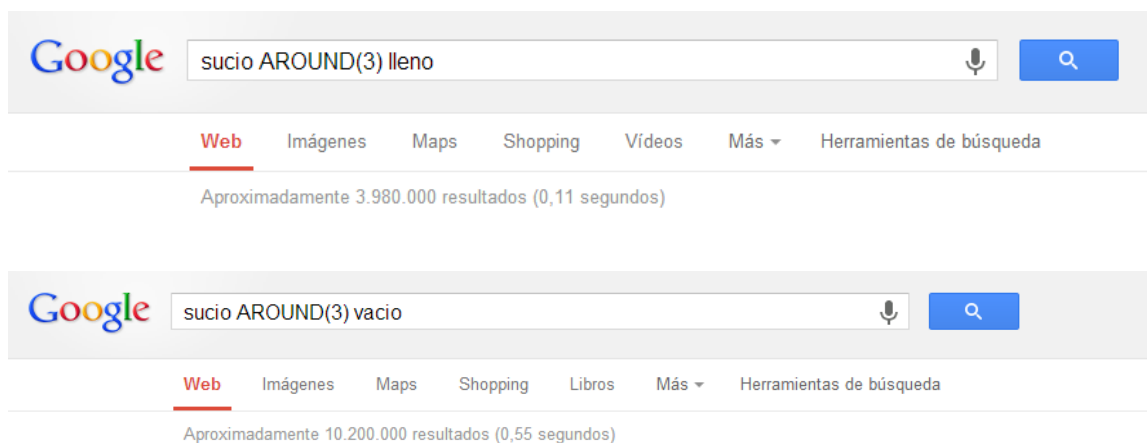


Ilustración 28. Resultado de la comparación sucio con lleno – vacío

Resultado para sucio – lleno = 3.980.000

Resultado para sucio – vacío = 10.200.000

Contador de positivos: 1 Contador de negativos: 3

Resultado provisional: negativo

Resultado final: Negativo



4.3 Prototipo de la interfaz

En esta parte del documento se va definir el prototipo que se ha diseñado para luego implementar la aplicación.

La aplicación únicamente constará de una pantalla que permitirá tanto introducir el nombre de la marca sobre la que se quiere obtener una valoración, como mostrar los resultados obtenidos una vez procesados los *tweets* en los que dicha marca es mencionada.

La **Ilustración 30** muestra el aspecto de dicha pantalla al arrancar la aplicación, mientras que en la figura 20 se muestra el aspecto que puede tener una vez el usuario a lanzado el proceso para una determinada marca concreta.

En este ejemplo concreto el idioma elegido es el español, por lo que toda la interfaz aparece en ese idioma.

La marca buscada ha sido “Coca-Cola”, y en el recuadro inferior se muestran los *tweets* recuperados, cada uno de ellos seguido de un color, verde, rojo y amarillo, según haya sido el resultado de sus procesamientos.

Por último, cabe reseñar las secciones “Sentimientos” y “Valoración Final”.

En la primera se muestra el número de *tweets* recuperados de cada sentimiento, mientras que la segunda se mostrará de forma global el “sentimiento” que tienen los usuarios de *Twitter* sobre esa marca.

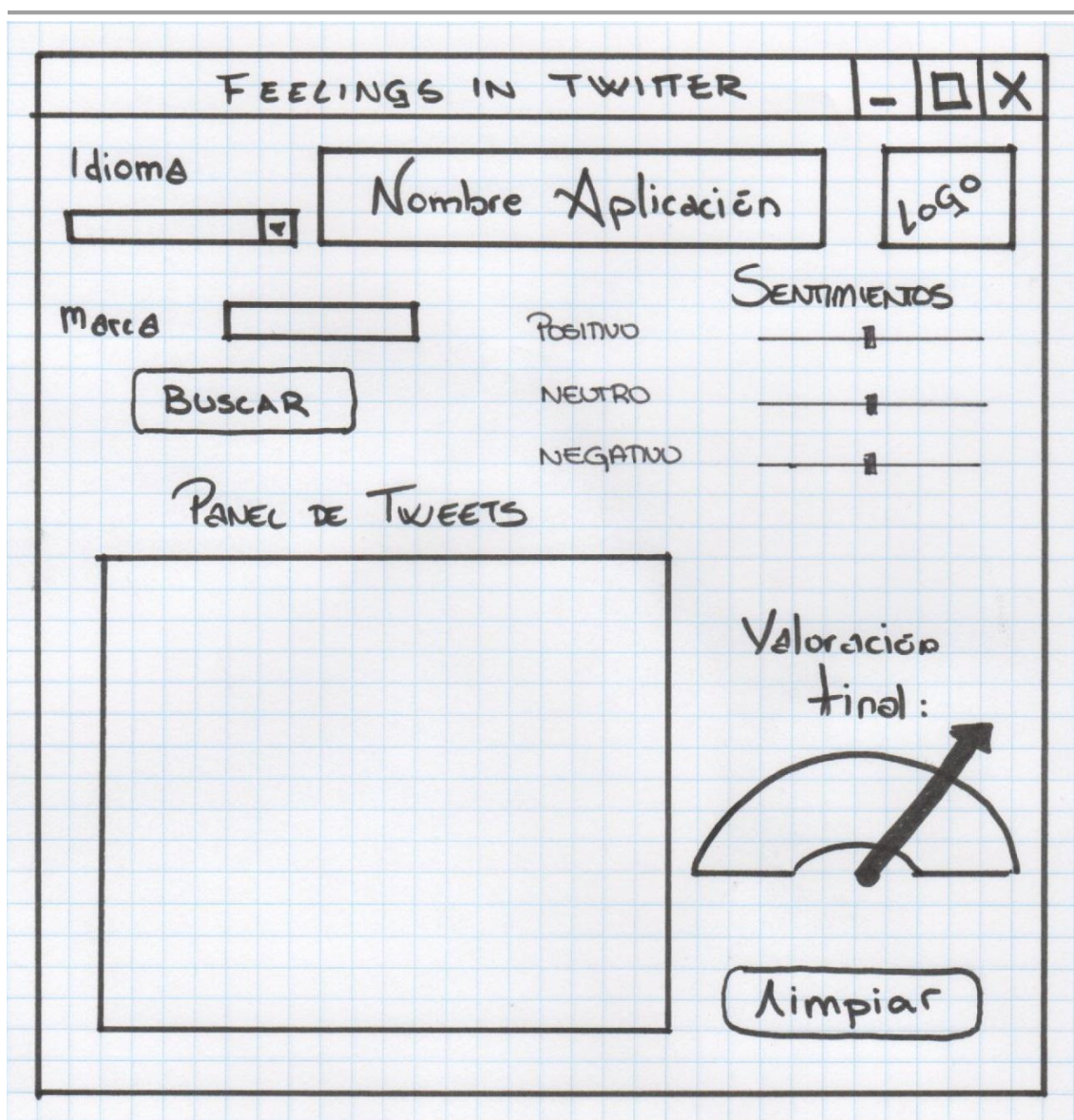


Ilustración 29. Prototipo de la pantalla inicial de la aplicación

FEELINGS IN TWITTER | - | □ | X

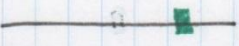
IDIOMA: **ESPAÑOL** ▼


Nombre Aplicación **Logo**


marca **Coca-Cola**

BUSCAR


SENTIMIENTOS


POSITIVO 


NEGATIVO 


NEUTRO 

PANEL DE TWEETS


TWEET 

TWEET 

TWEET 

TWEET 

Valoración final:



LIMPIAR

Ilustración 30. Prototipo de la pantalla inicial después de la búsqueda

4.4 Arquitectura del Sistema

4.4.1 Arquitectura de 3 capas

En este apartado se comentará la arquitectura que se ha elegido para la aplicación que se quiere desarrollar, en este caso se ha elegido la arquitectura en 3 capas, cuyo esquema se puede ver en la **Ilustración 31**.

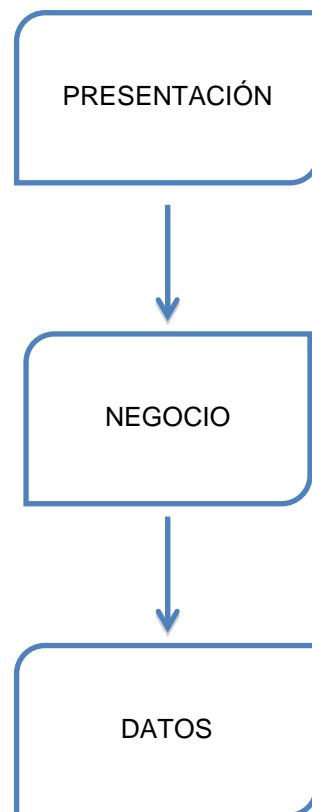


Ilustración 31. Modelo arquitectura de 3 capas

Tal y como se muestra en la **Ilustración 31** el patrón distingue entre 3 tipos de componentes fundamentales:

Presentación: en esta capa se encuentra la lógica que se necesita para la interacción entre usuario y la aplicación. Se puede dar a través de líneas de comando, sistemas de menús basados en textos o una interfaz de usuario.

Negocio: tiene que ver con la comunicación que se tiene entre otros sistemas en nombre de la propia aplicación, para realizar diferentes tareas. Esto puede ser API's, bases de datos o monitores de transacciones.

Datos: Lo que la aplicación realiza para su funcionamiento. Cálculos sobre las entradas, datos almacenados, validaciones.

Entre las principales ventajas que ofrece el patrón cabe destacar:

- Fácil de explicar y entender.
- Reduce dependencias y potencia la reutilización (cambiar interfaz sin tocar el resto).
- Permite dividir el trabajo en base a distintos roles (el diseñador gráfico no tiene por qué saber cómo se ha implementado el resto de componentes).

En la **Ilustración 32** se muestra la arquitectura de 3 capas, adaptado para el proyecto que se está desarrollando.

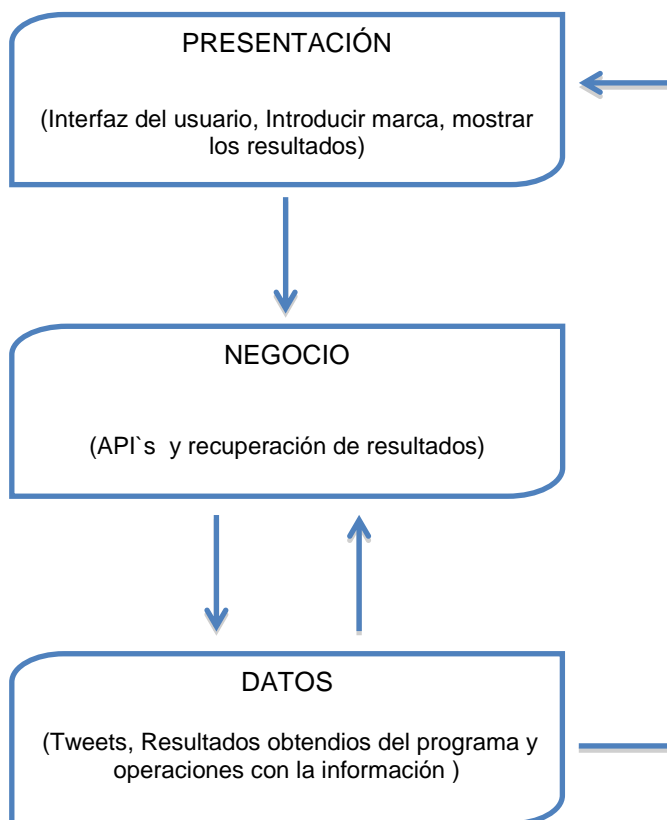


Ilustración 32. Modelo arquitectura de 3 capas para el proyecto

La **PRESENTACIÓN** es la parte que hace referencia a la interfaz, en la que va a comenzar todo el proceso y en la que también finalizará.

En la **PRESENTACIÓN** se tiene como punto de partida una caja de texto en la que se introducirá la marca a analizar. Al iniciar el proceso de búsqueda mediante el botón “Buscar”, éste lanzará una llamada a **NEGOCIO** que se encargará de buscar todos los tweets que contengan esa marca.



Un vez que negocio haya encontrado todos los tweets se los pasará a la capa de DATOS para que se realicen las operaciones necesarias para dar sentimiento a esos tweets.

Puntualmente se realizará una llamada a la capa de NEGOCIO desde la capa de DATOS para analizar los adjetivos que no se encuentren en la capa de DATOS y se necesite el API de google para analizarlos.

Una vez que se hayan analizado todos los tweets en la capa de DATOS, los resultados obtenidos se mostrarán en la caja de texto que se encuentra en la interfaz que está en la capa de PRESENTACIÓN.

4.4.2 Diagrama de clases

En este apartado se comentarán de qué forma están relacionadas las clases de la aplicación, cuáles son sus atributos y que métodos se pueden utilizar en cada una de ellas. En la **Ilustración 33** se muestra el diagrama global del proyecto.

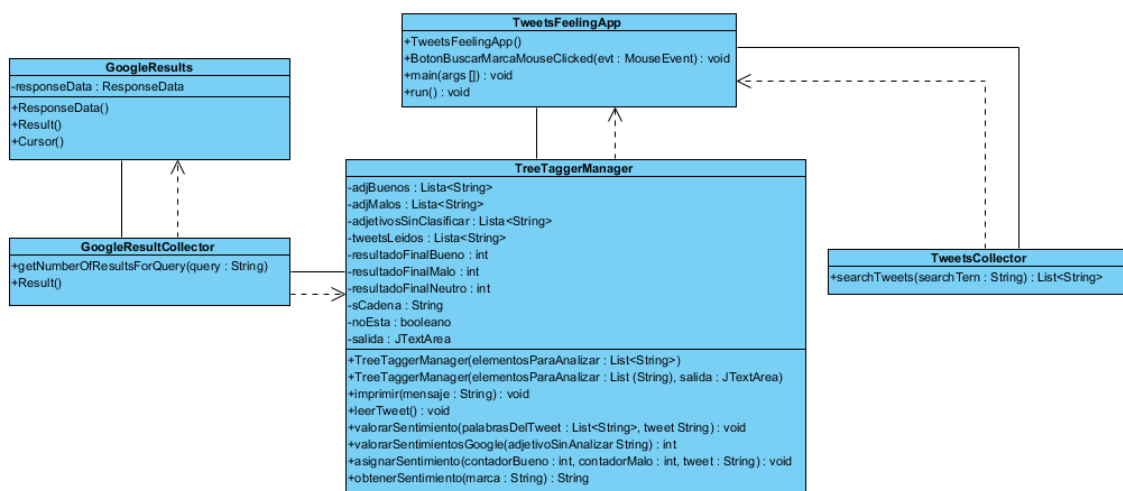


Ilustración 33. Diagrama de clases

Para poder explicar con mayor claridad el diagrama se va a dividir en 3 partes para explicar por qué se relacionan de esa forma las clases y ver con mayor claridad los atributos y métodos de cada una.

La primera relación entre clases, se puede observar en la **Ilustración 34**, es la que corresponde con la parte del API de Google. Este módulo hace referencia dentro de la arquitectura del sistema a la NEGOCIO.

En este caso se tienen dos clases *GoogleResults* y *GoogleResultColector*. Las cuales son necesarias si durante el análisis de los tweets, aparece un adjetivo sin clasificar.

Si ocurre esta situación, se generará una *query* que se le enviará a *GoogleResultColector*. Esta *query* se generará automáticamente y se irá modificando puesto que el adjetivo a clasificar irá cambiando.

Al método *GoogleResultColector* se le llamará desde la clase *TreeTaggerManager* dos veces, una para que compare con el *array* de buenos y otra con el *array* de malos, en cada llamada *GoogleResultColector* devolverá un valor numérico que será el que se evalúe para saber cómo clasificar al adjetivo.

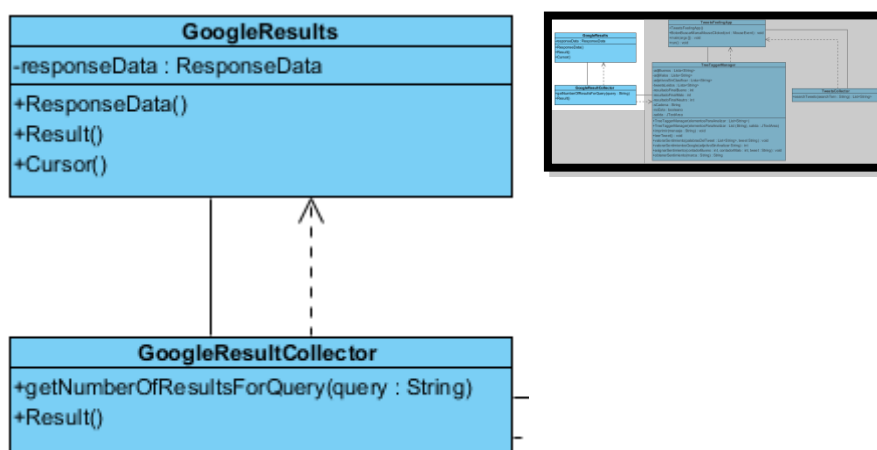


Ilustración 34. Diagrama de clases – Google

La segunda relación entre clases se ve reflejada en la **Ilustración 35** en este caso se observa que la relación es entre la clase principal *TweetsFeelingsApp* y *TweetsColleccto*. *TweetsFeelingsApp*, es el *main* de la aplicación, le pide a *TweetsColleccto*, clase del API de Twitter, que le pase todos los *tweets* que contenga la marca, la cual ha sido introducida por la interfaz. Este módulo hace referencia dentro de la arquitectura del sistema a la NEGOCIO.

El método en concreto que llama a *TweetsCollector* está dentro del método “*BotonBuscarMarcaMouseClicked*”. Esto quiere decir que en el momento en el que se pulse sobre el botón “Buscar” se ejecutarán todas las llamadas a métodos que estén programadas dentro de “*BotonBuscarMarcaMouseClicked*”.

Una vez se haya realizado la llamada, *TweetsCollector*, devolverá un objeto con todos los *tweets* que se han encontrado con esa marca. Por otro lado en *TweetsCollector* ya se ha indicado que se quieren en español y el número de *tweets* que se quieren. El API como máximo va a devolver 100 pero no quiere decir que siempre devuelva esa cantidad, puesto que no se tienen porque tener tantos *tweets* sobre esa marca, pero nunca devolverá más de 100.

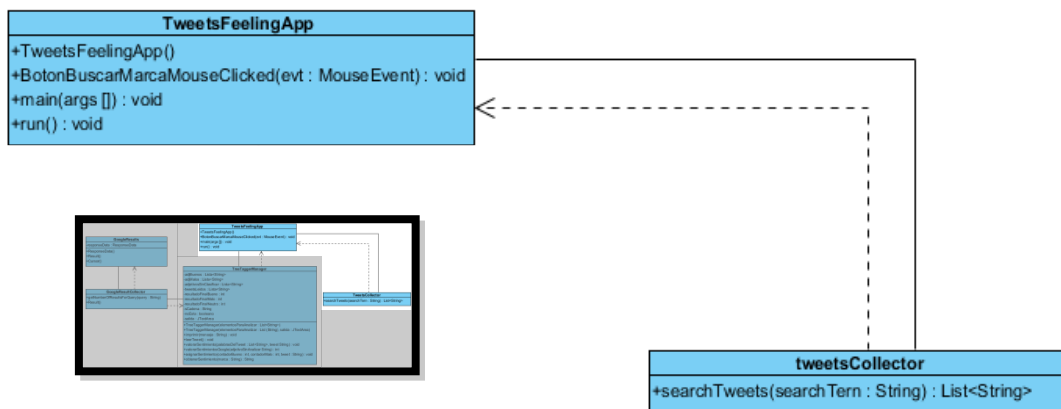


Ilustración 35. Diagrama de clases – Twitter

La tercera y última relación se puede observar en la [Ilustración 36](#), que hace referencia a la relación entre las 3 clases principales de la aplicación. Este módulo hace referencia dentro de la arquitectura del sistema a la PRESENTACION y DATOS.

Como ya se ha comentado *TweetsFeelingsApp* es la clase principal del proyecto, es la que lanza la interfaz y la que ejecuta el resto de métodos una vez se haya introducido un nombre de marca y presionado el botón “Buscar”.

A pesar de ser la clase principal, la clase que hace todo el trabajo computacional es *TreeTaggerManager*, el método que la inicia es *leerTweet()* que se encuentra dentro de “*BotonBuscarMarcaMouseClicked*” una vez llamado a ese método toda la ejecución pasa y se queda en *TreeTaggerManager* y es esta clase la que se encarga de llamar a *GoogleResults* si fuera necesario.

Como ya se ha comentado *GoogleResults* se encarga de realizar las comparaciones con el adjetivo a analizar. Una vez se tenga el resultado del adjetivo se devolverá a *TreeTaggerManager*.

Una vez que se ha acabado el análisis de todos los *tweets*, el resultado se retornará a *TweetsFeelingsApp* para que lo muestre por pantalla al usuario.

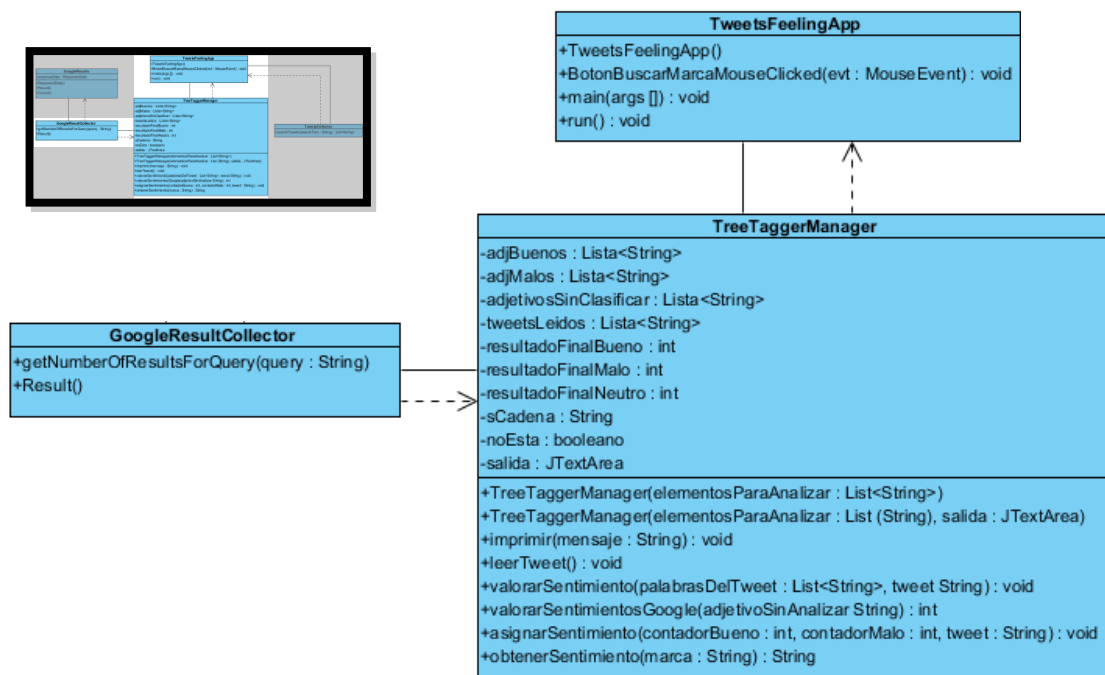


Ilustración 36. Diagrama de clases – Análisis de sentimientos acerca de marcas mediante Tweets

4.4.3 Diagrama de secuencia

En este apartado se explicará el diagrama de secuencia de la aplicación, en la **Ilustración 37** se puede observar que ejecución interna sigue el programa para realizar su función.

En el diagrama se detalla que métodos se utilizan en cada llamada, los atributos que necesitan para que se lleve a cabo el método, que retorna la llamada al método y como lo métodos se utilizan para realizar las llamadas entre clases.

Al igual que se hizo en el diagrama de clases, en este caso también se dividirá en tres para explicar cada uno de los métodos utilizados y ver con mayor claridad las relaciones entre las diferentes clases.

En el diagrama se muestran los nombres de las 4 clases más importantes que juegan un papel dentro de la ejecución de la aplicación.

Al igual que se mencionó en los diagramas de clases, en este caso *TweetsFellingsApp* es la clase principal, se ha situado en segundo lugar para evitar cruces de líneas y así dejar el diagrama más claro.

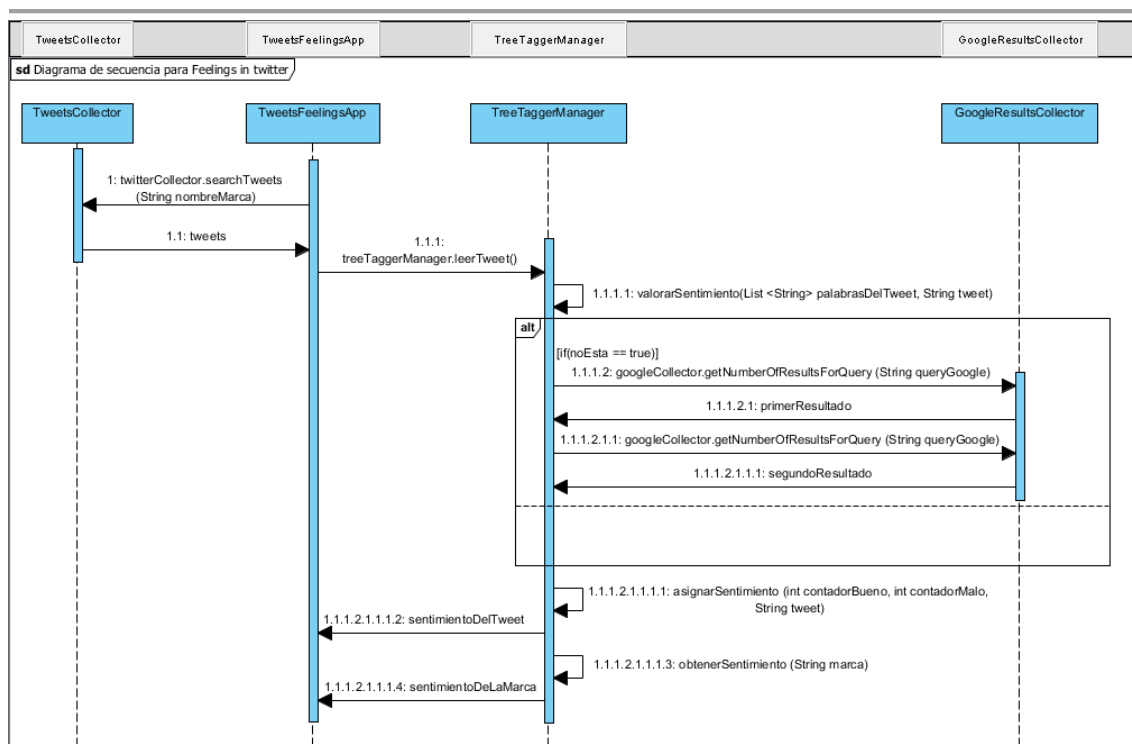


Ilustración 37. Diagrama de secuencia

El primer diagrama de secuencia, **Ilustración 38**, se observa la interacción entre la clase *TweetsCollector* y la clase *TweetsFeelingsApp* haciendo ésta de clase principal.

1. *TwitterCollector.searchTweets* (String nombreMarca). Este metodo se ejecuta de la clase *TweetsFeelingsApp* al introducir el nombre de la marca, que será el *string* que se le pasa al método, y pulsando el botón buscar.

1.1 *Tweets*. Son todos los tweets que devuelve la clase *TweetsCollector* a la clase *TweetsFeelingsApp* que contienen la marca que se ha pasado por parámetro. Esta clase es la que conecta con el API de *twitter* para conseguirlos.

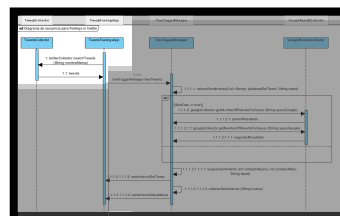
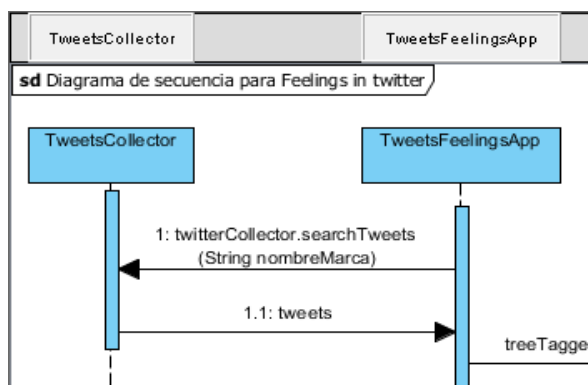




Ilustración 38. Diagrama de secuencia. Twitter – Análisis de sentimientos acerca de marcas mediante Tweets

El segundo diagrama de secuencia, **Ilustración 39** se observa la interacción entre la clase *TweetsFeelingsApp* (la clase principal) y la clase *TreeTaggerManager* es en la que se va a centrar el cómputo de la aplicación.

1.1.1. *TreeTaggerManager.LeerTweet()* se encarga de leer cada tweet que le llega de la aplicación *TweetsFeelingsApp*, una vez que tiene el tweet lo va a dividirlo por palabras, para que a continuación se guarden en un nuevo *array* únicamente las palabras, con esto se eliminan todos signos ortográficos de puntuación.

1.1.2.1.1.2. *SentimientoDelTweet*: este proceso se va a realizar tantas veces como tweets se hayan analizado. Este resultado se obtendrá una vez que la clase *TweetsFeelingsApp* ha realizado todos sus métodos correspondientes para obtener el análisis de los *tweets*.

1.1.2.1.1.4. *SentimientoDeLaMarca*: este proceso solo se va a realizar una única vez, por búsqueda. Una vez que estén analizados todos los *tweets* se mostrará con este proceso la decisión final que se tiene sobre la marca.

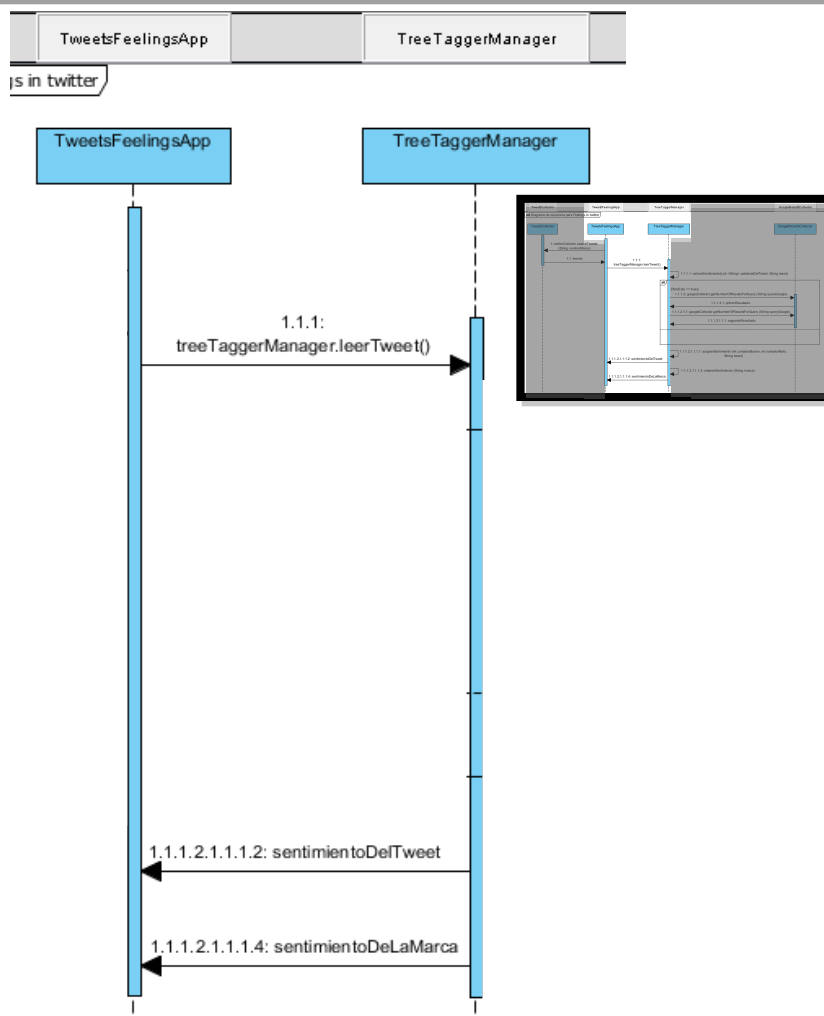


Ilustración 39. Diagrama de secuencia. Análisis de sentimientos acerca de marcas mediante Tweets - TreeTagger

El tercer diagrama de secuencia, **Ilustración 40**, se observa la interacción entre la clase *TreeTaggerManager* y la clase *GoogleResultsCollector*.

1.1.1.1 ValorarSentimiento (List <String> palabrasDelTweet, String tweet) se le pasa el *array* que se consiguió en el punto 1.1.1 junto con todo el *tweet* sin separar, tal y como se consiguió del *API* de *twitter*.

Los siguientes procesos se ejecutarán si alguno de los adjetivos que se están analizando en *ValorarSentimiento()* no se han podido analizar en ese método. Saber si se ha analizado o no un adjetivo se sabe gracias al *booleano* *noEsta*, que es el que se comparará para entrar a ejecutar los siguientes procesos.

1.1.1.2 `googleCollector.getNumberOfResultsForQuery` (String queryGoogle). Este proceso recibe un *queryGoogle*, que se realiza en el



momento de enviárselo a *google*, en este caso sería el adjetivo a analizar más AROUND (3) más el adjetivo correspondiente al *array* de buenos.

1.1.1.2.1 primerResultado. Devuelve el resultado de ejecutar la *query* en el *API* de *Google*, este resultado se guardará para comprobarlo con el siguiente resultado, segundoResultado.

1.1.1.2.1.1 *googleCollector.getNumberOfResultsForQuery* (String *queryGoogle*) Este proceso recibe un *queryGoogle*, que se realiza en el momento de enviárselo a *google*, en este caso sería el adjetivo a analizar mas AROUND (3) mas el adjetivo correspondiente al *array* de malos.

1.1.1.2.1.1.1 segundoResultado. Devuelve el resultado de ejecutar la *query* en *google*, este resultado se guardará y comparará con el resultado anterior, primer resultado.

De la comparación de ambos resultado se sacará el análisis de ese adjetivo y se tendrá en cuenta para evaluarlo. Una vez que se hayan analizado todos los adjetivos se pasarán a los siguientes procesos.

Los siguientes procesos se ejecutarán siempre, indistintamente de que se entre o no en la condición *if*.

1.1.1.2.1.1.1.1 *asignarSentimiento* (int *contadorBueno*, int *contadorMalo*, String *tweet*). Durante la ejecución de *ValorarSentimiento()* en este caso se dará el sentimiento de forma individual al *tweet* gracias a los valores de los contadores, que son el número de adjetivos positivo o negativos que se han encontrado durante la evaluación del *tweet* y sus adjetivos. Este método se ejecutará tantas veces como *tweets* se hayan analizado.

1.1.1.2.1.1.1.3 *obtenerSentimiento* (String *marca*) este proceso solo se ejecutará una vez por búsqueda, esto se así puesto que solo se da una vez la valoración de la marca y es justamente al final del análisis de todos los *tweets* que se han encontrado sobre esa marca.

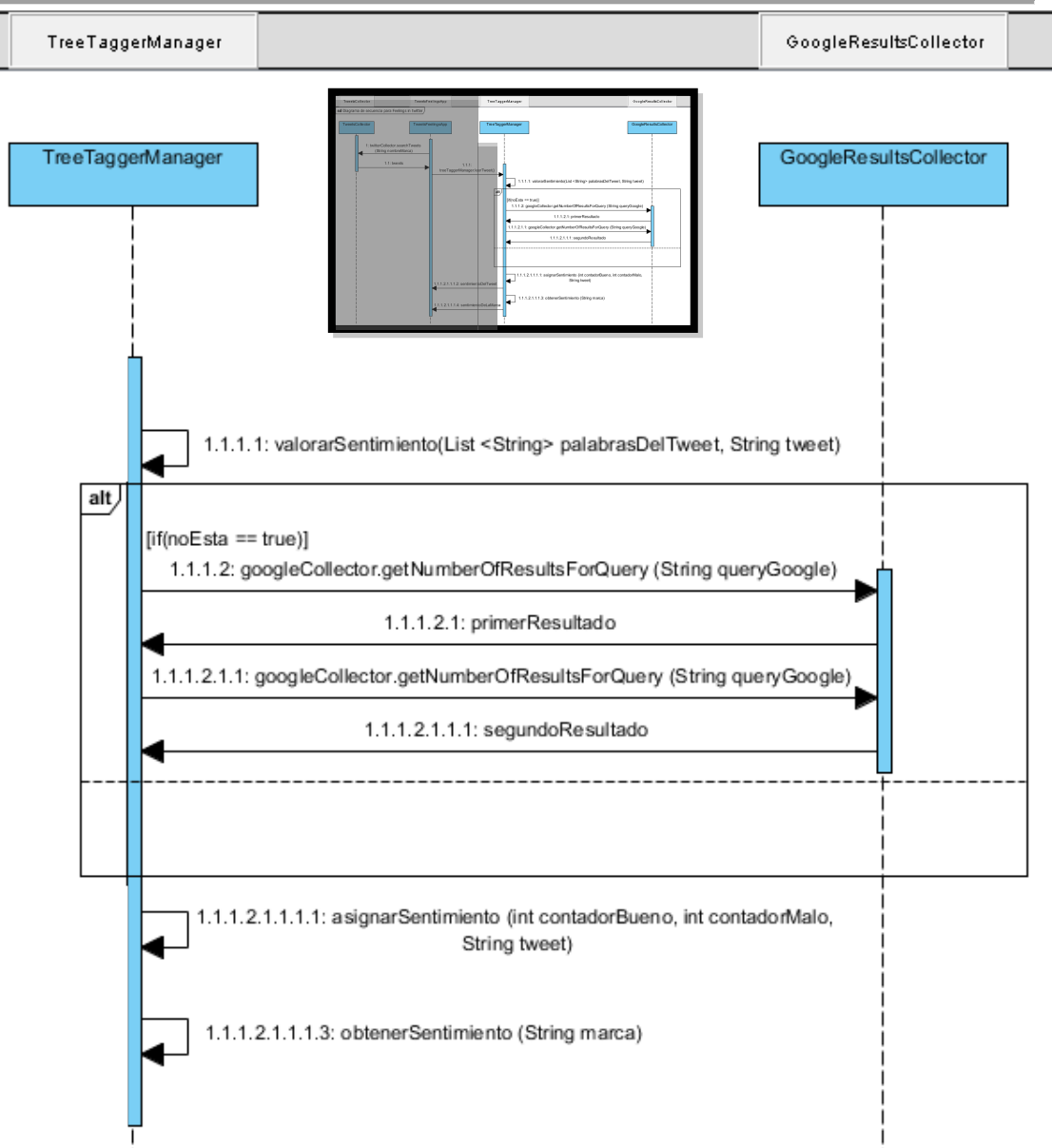


Ilustración 40. Diagrama de secuencia. TreeTagger - Google



CAPÍTULO 5

Implementación y Desarrollo

En este capítulo se va a mostrar la aplicación final que se va a entregar al usuario.

Se explicarán las tecnologías utilizadas para generar las interfaces y la aplicación en sí.

Se van a mostrar dos interfaces, la primera mostrará la pantalla antes de realizar ninguna búsqueda, y la segunda los resultados obtenidos una vez realizada la búsqueda de una marca.





5 IMPLEMENTACIÓN Y DESARROLLO

A partir del prototipo que se presenta en el punto 4.3 se ha realizado un prototipo de alto nivel. Se va a mostrar la implementación que se ha diseñado para este proyecto.

5.1 Tecnologías utilizadas

Para poder comentar las tecnologías utilizadas hay que dividir este apartado en dos, uno para comentar la tecnología relacionada con la **programación de la aplicación**, y la otra la utilizada para **mostrar la aplicación**.

Para la **programación de la aplicación** se ha utilizado como herramienta *NetBeans* y como lenguaje java. Se ha decidido hacerlo de esta forma puesto que la aplicación quiere ser portable.

Por otro lado java es el mejor lenguaje de programación para poder integrar la aplicación realizada con internet y con la red social twitter. Por otro lado java es portable y con una modificación mínima en el código se podría cambiar e introducir cualquier otra red social o aplicación.

Para mostrar la aplicación este tema es un tanto complejo, hoy en día existen gran cantidad de dispositivos electrónicos, la aplicación va a poder ser utilizada en cualquiera de ellos, por lo que se ha tendido que tener en cuenta las características que tienen cada uno de ellos.

La mayor característica a la que se ha tenido que hacer frente es la del tamaño de pantalla, esto es un tema muy importante, puesto que no se tiene el mismo tamaño en un pc que en un portátil o entre los diferentes tipos de tablets o smartphones, por lo que se ha diseñado de tal forma la aplicación que se pueda ajustar automáticamente a cualquier tipo de tamaño de pantalla.

Otra característica puede ser los diferentes tipos de sistemas operativos que se pueden encontrar en el mercado, esta característica como tal no nos afecta puesto que la aplicación va a ser utilizada a través de Internet pero si se ha tenido en cuenta que funcione en todos los navegadores y todas sus versiones.

5.2 Interfaz

Para crear la interfaz se ha seguido el prototipo, que se presentó en el punto 4.3, como se verá a continuación, se han tenido que realizar algunos cambios con respecto al prototipo.



Los cambios que se enumeran a continuación, se pueden observar en las siguientes ilustraciones: **Ilustración 41**, **Ilustración 42**, **Ilustración 43**.

- Aumento considerable de la letra de los textos de la interfaz.
- Cambio en los colores de los textos, para evitar que personas con problemas visuales puedan malinterpretarlos.
- Nuevas barras de progresión para dar más información sobre los resultados de los sentimientos individuales.
- Cambio de la barra para el estado global de la marca.
- Nuevo texto para aportar más información para el significado global de la marca.
- La interfaz, solo se puede mostrar en un idioma.

En la **Ilustración 41** se muestra como es la interfaz una vez que se el usuario acceda a la misma. Todas las barras tienen que estar en su posición inicial (0) y el estado de proceso “*En espera*”.

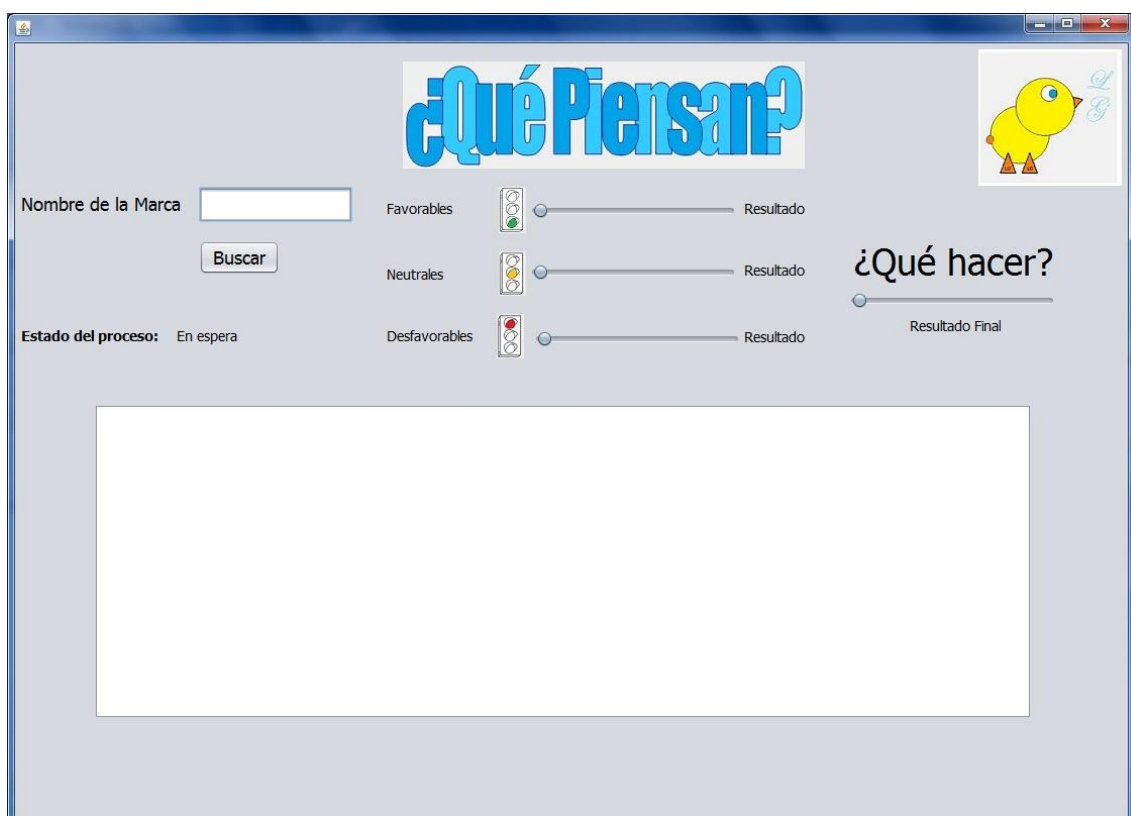


Ilustración 41. Interfaz de la aplicación antes de iniciar una búsqueda



Análisis de sentimientos acerca de marcas mediante Tweets

En la **Ilustración 42** se muestra que componentes de la interfaz cambian, en el momento en el que el usuario introduce una marca a analizar.

La caja de texto, pasa a tener el nombre de la marca a analizar.

El botón buscar cambia ligeramente de color.

El estado del proceso pasa a estar “En ejecución”.

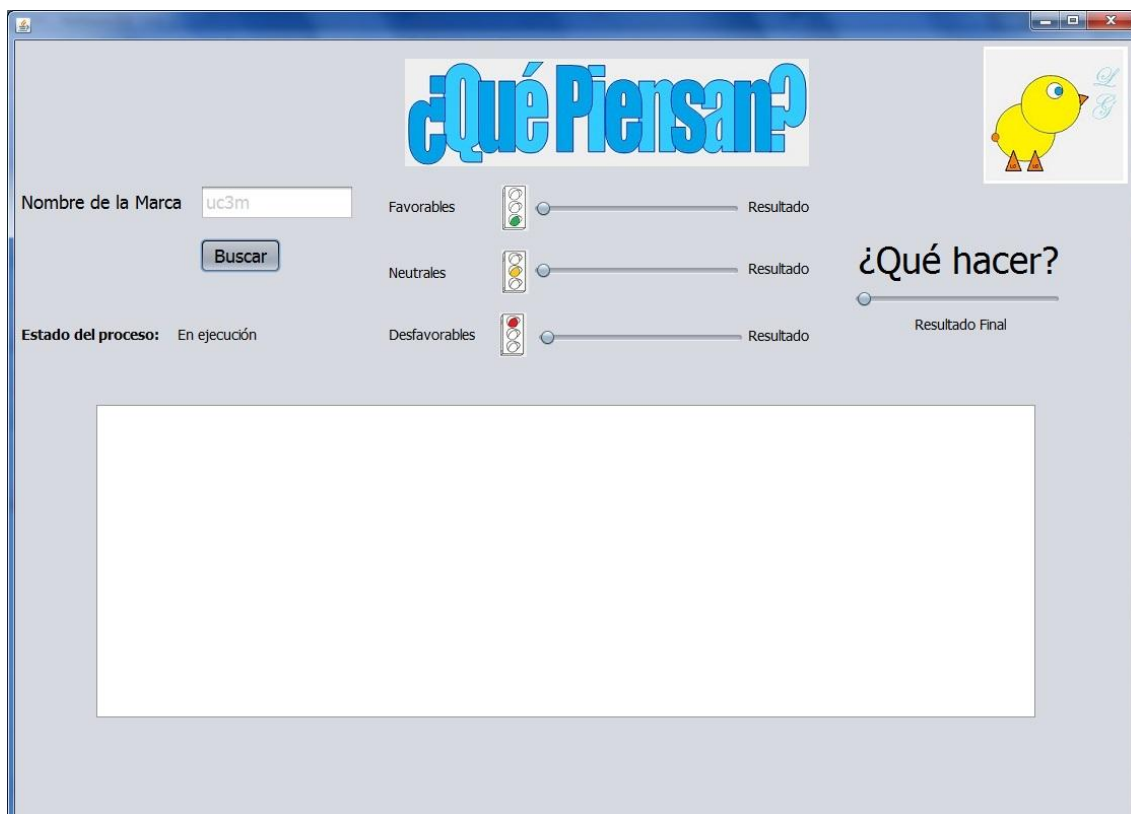


Ilustración 42. Interfaz de la aplicación al introducir una marca

En la **Ilustración 43** se muestran que muestra que componentes de la interfaz cambian, en el momento en el que la interfaz acaba de analizar todos los tweets.

El botón buscar vuelve a la tonalidad inicial.

El estado del proceso cambia a “Finalizado”.

Las barras se mueven en función de los resultados obtenidos.

Los textos “Resultado” y “Resultado final” cambiarán por números, indicando el número de tweets de ese sentimiento y por una frase explicativa de la marca global, respectivamente.



Ilustración 43. Interfaz de la aplicación después de iniciar una búsqueda



CAPÍTULO 6

Gestión del Proyecto

En esta sección, el documento mostrará todo lo relacionado con el ciclo de vida que se ha utilizado, la planificación que se ha diseñado, los detalles económicos correspondientes a la presente oferta, así como la forma de pago, definiendo los términos de la garantía, los cuales ofrece la aplicación.



6 GESTIÓN DEL PROYECTO

6.1 Modelo del ciclo de vida del software

La planificación está dividida en diversos apartados, los cuales corresponden a las fases del modelo de ciclo de vida en cascada.

Se va a utilizar este modelo por tres motivos, el primero es por cómo se ha diseñado la planificación, el segundo puesto que al no ser un proyecto de gran envergadura es un buen modelo a utilizar y por último, al ser un proyecto bien definido de principio a fin cumple con las condiciones básicas para poder utilizarlo, en la **Ilustración 44** se muestran de forma gráfica.



Ilustración 44. Ciclo de vida en cascada

En el momento en el que se termine una fase y se pase a la siguiente, se volverá al anterior, para afianzar que todo lo realizado en la fase anterior es correcto y concuerda con la siguiente fase.



6.2 Planificación

En esta parte del documento se va a mostrar en la [Ilustración 45](#) la planificación del proyecto mediante el diagrama de Gantt. En la [Tabla 33](#), se mostrará de una forma más detallada las fases y subfases con su duración y fecha de inicio y fecha de fin.

La planificación se divide principalmente en dos entregas, la primera de ellas se corresponde con el anteproyecto del sistema (se corresponde con el primer hito) y una segunda parte que se corresponde con la entrega del proyecto completo (se corresponde con el segundo hito).

Para la planificación se ha obtenido un número total de horas de **1344 horas**, este número de horas ha sido calculado en base al plazo que se dispone para la realización del proyecto completo que es de **6 meses**.

El proyecto se ha dividido en 6 fases: planificación, investigación (se divide en dos partes), análisis, diseño, implementación y documentación.

Alguna de estas fases ha sido realizada en paralelo puesto que era más eficiente para no olvidar nada y la parte de investigación se ha dividido, puesto que la información a buscar tenía diferentes objetivos según el momento en el que se encontrara el proyecto.

A continuación se explicara en qué consiste cada una de las fases.

- **Planificación:** esta es la fase más importante del proyecto puesto que es en la que se va a reflejar de qué forma se va a dividir el trabajo y el tiempo estimado que se va a dedicar a cada fase.
- **Investigación:** esta parte es importante, puesto que se realiza una primera búsqueda para tener una idea sobre el tema que se va a desarrollar, saber si ya se ha creado, si es útil o si tiene futuro entre otras cuestiones.
- **Análisis:** en esta fase se definirán los requisitos del sistema que se va a desarrollar, incluyendo los requisitos de usuarios, software y hardware.
- **Diseño:** es la parte que se dedica tiempo a la creación del prototipo cubriendo los requisitos mostrados en el análisis, explicando que tecnologías se han utilizado.
- **Implementación:** en esta fase se incluye mayoritariamente la codificación de los componentes detallados en la fase anterior



comentando en que sistemas y dispositivos se puede utilizar la aplicación creada.

- **Investigación:** se tiene que realizar una nueva investigación para profundizar en todos los conceptos que se encontraron en la primera investigación y así poder desarrollarlos.
- **Documentación:** redactar y esquematizar todo lo que se ha encontrado en la segunda investigación y plasmarlo en el presente documento, para explicar el proyecto realizado.

Fase	Duración	Fecha de inicio	Fecha de fin
Investigación	9 days	Sun 14/04/13	Wed 24/04/13
Búsqueda Inicial	9 days	Sun 14/04/13	Wed 24/04/13
Análisis	3 days	Thu 25/04/13	Sun 28/04/13
Requisitos	2 days	Thu 25/04/13	Fri 26/04/13
Casos de Uso	2 days	Fri 26/04/13	Sun 28/04/13
Diseño	10 days	Sun 28/04/13	Thu 09/05/13
Prototipo	3 days	Sun 28/04/13	Tue 30/04/13
Aplicación	6 days	Tue 30/04/13	Tue 07/05/13
Tecnologías Aplicadas	3 days	Tue 07/05/13	Thu 09/05/13
Implementación	38 days	Thu 09/05/13	Mon 01/07/13
Aplicación	29 days	Thu 09/05/13	Tue 18/06/13
Interfaz	10 days	Tue 18/06/13	Mon 01/07/13
Investigación	45 days	Mon 01/07/13	Thu 29/08/13
Contexto	11 days	Mon 01/07/13	Mon 15/07/13
Twitter	7 days	Mon 01/07/13	Tue 09/07/13
Tuenti	3 days	Tue 09/07/13	Thu 11/07/13
Facebook	3 days	Thu 11/07/13	Mon 15/07/13
Estado del Arte	35 days	Mon 15/07/13	Thu 29/08/13
Análisis de Sentimientos	14 days	Mon 15/07/13	Thu 01/08/13
Twitter	4 days	Thu 01/08/13	Tue 06/08/13
TreeTagger	13 days	Tue 06/08/13	Thu 22/08/13
Google y Bing	7 days	Thu 22/08/13	Thu 29/08/13
Documentación	43 days	Mon 01/07/13	Tue 27/08/13
Introducción	9 days	Mon 01/07/13	Thu 11/07/13
Estado del Arte	14 days	Thu 11/07/13	Tue 30/07/13
Análisis	4 days	Tue 30/07/13	Fri 02/08/13
Diseño	4 days	Fri 02/08/13	Wed 07/08/13
Implementación y Desarrollo	1 day	Wed 07/08/13	Wed 07/08/13
Gestión del Proyecto	4 days	Thu 08/08/13	Tue 13/08/13
Conclusiones	6 days	Tue 13/08/13	Tue 20/08/13
Trabajo Futuro	6 days	Tue 20/08/13	Mon 26/08/13
Presentación	10 days	Mon 02/09/13	Fri 13/09/13
Revisión del Proyecto	14 days	Fri 13/09/13	Wed 02/10/13
Presentación del Proyecto	1 day	Mon 04/11/13	Mon 04/11/13

Tabla 33. Planificación del Proyecto

Tras mostrar las fases en las que se han dividido la planificación del proyecto, se muestran éstas dentro de un diagrama de Gantt donde se



Análisis de sentimientos acerca de marcas mediante Tweets

muestran de manera temporal cada una de las tareas con sus correspondientes dependencias.

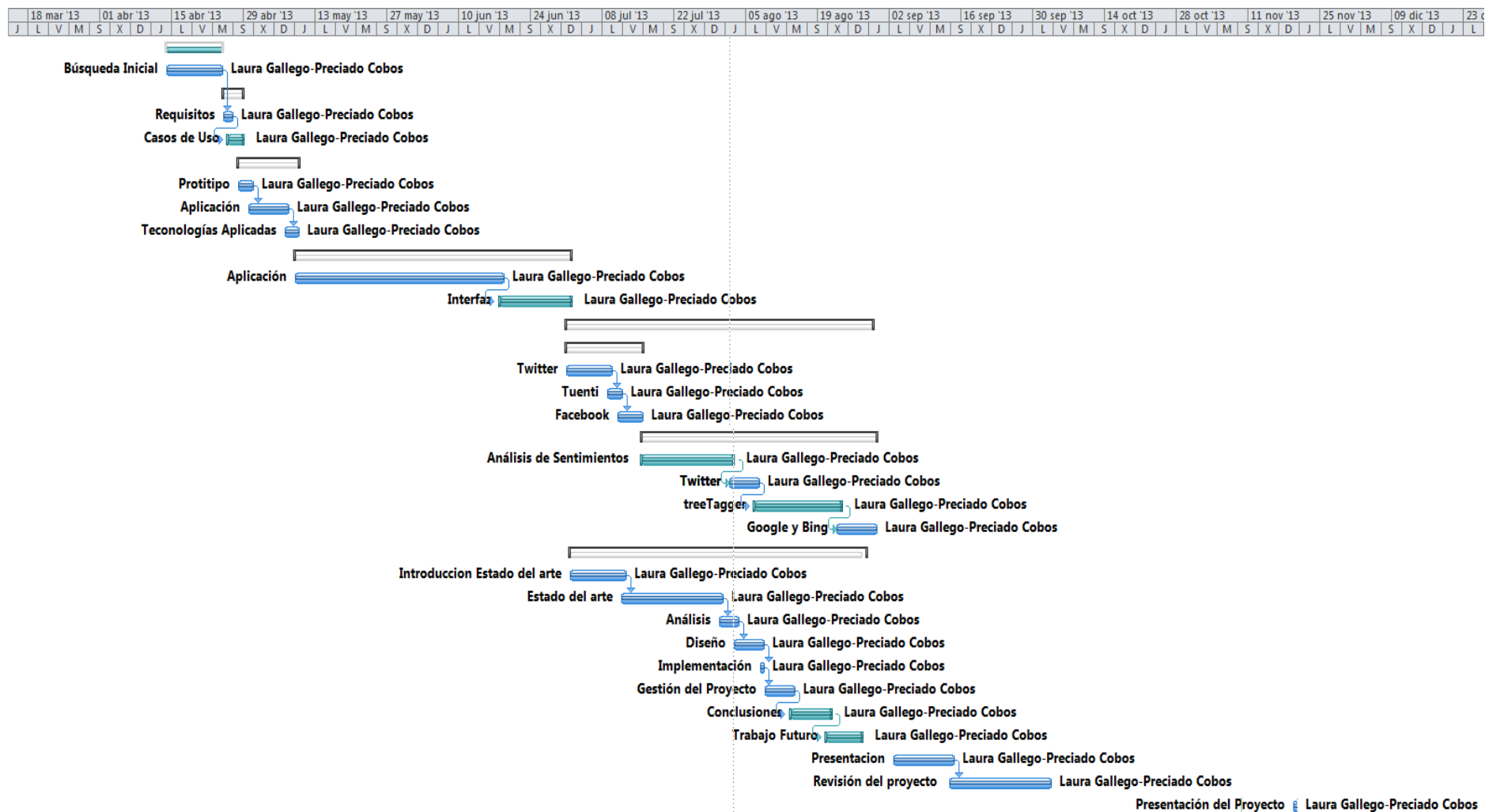


Ilustración 45. Diagrama de Gantt



6.3 Presupuesto

6.3.1 Oferta

La oferta que se presenta del proyecto “Análisis de sentimientos acerca de marcas mediante Tweets”, se valora el precio del software a realizar en **137.132.87 €** IVA del 21% incluido.

6.3.2 Periodo de validez

La validez de la presente oferta se fijará en **VEINTE** días hábiles a partir de la fecha de la entrega de este presupuesto, día **15 de septiembre de 2013**.

6.3.3 Costes

En este apartado se detallará el presupuesto que cubra todos los costes y las necesidades que se irán planteando a lo largo de la evolución del proyecto, que están relacionados con éste.

En la especificación de los siguientes costes, se ha incluido cierta cantidad en concepto de hardware y software utilizado por el personal de la empresa, en dicho caso se ha atribuido una cantidad correspondiente a la amortización del proyecto.

Esta amortización se realiza a un periodo de 24 meses ya que es la que se suele utilizar en los productos adquiridos por una empresa, como es este caso.

6.3.3.1 Coste personal

En este apartado se va a mostrar los costes de personal, de los integrantes que van a desarrollar el proyecto “Análisis de sentimientos acerca de marcas mediante Tweets”.

El proyecto se va a desarrollar por una única persona, la cual va a ser la encargada de realizar diferentes cargos a lo largo del documento. En la **Tabla 34** se mostrará el coste/hora y los diferentes cargos que va representar dicha persona.



Análisis de sentimientos acerca de marcas mediante Tweets

Categoría	Coste / Hora (€)
Jefe de proyecto/ Analista / Programador / Diseñador	60.00 €

Tabla 34. Coste por Hora/Categoría

Por la planificación mostrada en el punto 6.1, en la **Tabla 35** se resume el coste de cada fase según el cargo. Se multiplicarán las horas por el coste de cada hora.

Cargo	Investigación	Planificación	Análisis	Coste por fase		Investigación	Documentación	Coste total (€)
				Diseño	Implementación			
Jefe de proyecto / Analista / Programador / Diseñador	70 X 60€ = 4,200,00 €	30 X 60€ = 1,800,00 €	83 X 60€ = 4,980,00 €	270 X 60€ = 16,200,00 €	380 X 60€ = 22,800,00 €	70 X 60€ = 20,400,00 €	100 X 60€ = 4,200,00 €	6,000,00 €

Tabla 35. Total Coste Personal

6.3.3.2 Coste equipos

En la **Tabla 36** se mostrarán el precio de los equipos informáticos que se necesitarán para llevar a cabo el proyecto.

Descripción	Coste / Unidad	Unid	Coste total	Periodo de Amortización (meses)	Duración proyecto (meses)	Coste asociado proyecto (€)
Impresora HP OFFICEJET 6000	54.95 €	1	54.95 €	24	6	13.74 €
Ordenador HP Compaq 610	225.35 €	1	225.35 €	24	6	56.34 €
HP Pavilion dm4	426.60 €	1	426.60 €	24	6	106.65 €
Logitech Ratón M125 – Blanco	10.96 €	1	10.96 €	24	6	2.74 €
Logitech Ratón inalámbrico M195	14.95 €	1	14.95 €	24	6	3.74 €
Total						183.20 €

Tabla 36. Total Coste de equipos

6.3.3.3 Coste software

En la **Tabla 37** se mostrarán los costes sufridos al adquirir las licencias para el software utilizados para desarrollar el proyecto.



Análisis de sentimientos acerca de marcas mediante Tweets

Producto	Coste / Unidad	Unid	Coste total	Periodo de Amortización (meses)	Duración proyecto (meses)	Coste asociado proyecto
Windows 7 Home Premium	349.00 €	1	349.00 €	24	6	87.25 €
Microsoft Office 2010	699.00 €	1	699.00 €	24	6	174.75 €
Office Project Standard 2010	775.00 €	1	775.00 €	24	6	193.75 €
Altova UModel 2009 Enterprise Edition	139.80 €	1	139.80 €	24	6	34.95 €
Total						490.70 €

Tabla 37. Total Coste de software

6.3.3.4 Coste material fungible

En la **Tabla 38** se mostrarán todo el material fungible que se necesita para poder desarrollar el proyecto.

Descripción	Precio/unidad	Unidad	Coste total (€)
Material de papelería			100.00 €
HP OFFICEJET 6000 NEGRO	9.99 €	1	9.99 €
HP OFFICEJET 6000 AMARILLO	7.69 €	1	7.69 €
HP OFFICEJET 6000 CYAN	7.69 €	1	7.69 €
HP OFFICEJET 6000 MAGENTA	7.69 €	1	7.69 €
Total			133.06 €

Tabla 38. Total Coste Material fungible

6.3.3.5 Coste viajes y dietas

En este apartado se contabilizará los gastos relativos a viajes y a las dietas de los miembros de la empresa producida en este proyecto.

Descripción	Coste / Km	Días / mes	Km / día	Meses	Vehículos	Coste (€)
Complemento gasolina	0.22 €	20	50	6	1	1.320.00 €

Tabla 39. Total Coste de viajes y dietas



Análisis de sentimientos acerca de marcas mediante Tweets

6.3.4 Costes Directos Totales

En la **Tabla 40** se resumirán todos los costes directos que se han asociado al proyecto por los conceptos de personal, gastos de software y gastos de equipos y de maquinaria para la planta de fabricación.

Descripción	Coste Total
Personal	80.640.00 €
Equipos y Dispositivos	183.20 €
Software	490.70 €
Material Fungible	133.06 €
Viajes y dietas	1.320.00 €
Total	82.766.96 €

Tabla 40. Total Costes directos

6.3.5 Costes Indirectos

En la **Tabla 41** se mostrarán los costes indirectos, para el proyecto se ha estimado que sea un 4,5%.

Gastos Totales	Margen de Costes Indirectos	Costes Indirectos
82.766.96 €	4.50%	3.724.51 €

Tabla 41. Total Costes indirectos

6.3.6 Beneficio

En la **Tabla 42** se presenta el beneficio que se va a obtener de realizar el sistema. El beneficio que se cree apropiado sería del 20%.

Gastos Totales	Margen de Costes Beneficios	Beneficio total
2.766.96 €	20%	16.553.39 €

Tabla 42. Beneficio



Análisis de sentimientos acerca de marcas mediante Tweets

6.3.7 Margen de Riesgo

En la **Tabla 43** se mostrará el margen de riesgo que se ha estimado en un 15%.

Gastos Totales	Margen riesgo	Total de riesgos
82.766.96 €	15%	12.415.04 €

Tabla 43. Margen de riesgo

6.3.8 Presupuesto Total

El presupuesto que se muestra en la **Tabla 44** se representa sin IVA. Por lo que al presupuesto final se debe incluir el 21% de IVA.

Descripción	Euros
Gastos Totales	80.640.00 €
Costes Indirectos	3.724.51 €
Beneficio	16.553.39 €
Riesgos	12.415.04 €
Total	113.332.95 €

Tabla 44. Presupuesto total sin IVA

Aplicando el correspondiente IVA del 21%, el presupuesto total se queda en:

Descripción	Valor	Porcentaje	Euros
Presupuesto con IVA	137.132.87 €	21%	23.799.92 €

Tabla 45. Presupuesto total con IVA

6.3.9 Forma de pago

El pago para la realización del proyecto se fraccionará en **TRES**, de acuerdo a las siguientes condiciones:



Análisis de sentimientos acerca de marcas mediante Tweets

- Abono de la primera cuota, con un valor igual al **25%** del precio total en el momento de la firma del contrato, el día **15 de junio de 2013**
- Abono de la segunda cuota, con valor igual al **35%** del precio total el día **1 de septiembre de 2013**.
- Abono de la tercera y última cuota, con un valor igual al **40%** del precio total el día **4 de noviembre de 2013**, coincidiendo con la entrega e implantación del sistema.

En la **Tabla 46** se muestra en detalle la cuantía que se debe abonar en cada uno de los pagos indicados con anterioridad. Los precios incluyen un 21% de IVA.

Cuota	Año fiscal	Importe	IVA %	IVA	Total a pagar
Primera cuota	2013	28.333.24 €	21.00%	5.949.98 €	34.283.22 €
Segunda cuota	2013	39.666.53 €	21.00%	8.329.97 €	47.996.50 €
Tercera cuota	2013	45.333.18 €	21.00%	9.519.97 €	54.853.15 €
Total					137.132.87 €

Tabla 46. Resumen de Forma de Pago



6.3.10 Garantía

El producto que se está ofertando, está cubierto por una garantía ante posibles defectos en su funcionalidad durante un periodo de **DIECIOCHO MESES** a partir de la fecha de puesta en marcha del mismo.

La garantía que se ofrece, incluye:

- Soporte telefónico o a través de Internet para solventar cualquier problema o necesidad que pueda surgir sobre el producto implantado.
- Asignación de profesionales cualificados para mantener el producto, solucionando fallos del producto que puedan producirse durante el funcionamiento del mismo.
- Tiempo de respuesta ante incidencias de cinco horas como máximo para aquellas que sean críticas para el adecuado funcionamiento del producto, que paralicen sus actividades y surjan en días laborables y tiempo de respuesta máximo de un día hábil para evaluar y proporcionar una estimación del tiempo de reparación de incidencias no críticas que pudieran aparecer.

Opcionalmente y por un precio a acordar se ofrece:

- Formación a nuevos usuarios del producto.
- Soporte 24/7 ante incidencias, con tiempos de respuesta de 5 horas para las incidencias críticas que paralicen sus actividades de negocio y 72 horas desde la comunicación de la incidencia hasta su resolución en el caso de las incidencias de gravedad baja que no paralicen su actividad.

La presente garantía no cubre pérdidas de información derivadas de un mal uso o manipulación indebida del producto u otros usos negligentes del mismo.

Una vez finalizado el periodo de garantía, se le ofrece poder negociar una extensión del periodo de soporte y garantía del producto por un precio adicional al que se incluye en este presupuesto.

6.3.10.1 Anulación de la garantía

Se considerará inválida la presente garantía por causas tales como:

- Uso de un entorno tecnológico diferente al que se negocie.



Análisis de sentimientos acerca de marcas mediante Tweets

- Usos negligentes del producto.
- Manipulación del producto por personal no autorizado.



CAPÍTULO 7

Conclusiones

En esta sección, se va a comentar las conclusiones a las que se han llegado al acabar el presente documento.



7 CONCLUSIONES

En este apartado se va a mostrar a que conclusiones a las que se han llegado durante la realización del proyecto, también se mostrarán las ideas que se tenían inicialmente sobre este proyecto y comentar si se han mantenido o modificado a lo largo de su desarrollo.

La primera idea inicial que se tuvo sobre el proyecto fue: “Interesante, pero... ¿Alguien lo utilizará?”. A medida que se ha ido leyendo la información, la cual se ha tenido que leer, procesar y recopilar para poder realizar este proyecto, se auto respondió de forma amplia y sorprendente la cuestión que se planteaba al principio del proyecto. Es impresionante, la cantidad de gente que está interesada en el tema del análisis de sentimientos, (como se ha querido plasmar a lo largo del documento, toda esta ciencia va sobre textos, opiniones y comentarios que se dejan de forma escrita en cualquier lugar, en este caso se ha centrado la atención en Internet) y la necesidad que tiene la gente en poder tratarlos, para saber que quieren decir. Llega a ser tan necesario que hay empresas especializadas en este tipo de cuestiones, las cuales ofrecen sus servicios a distintas empresas o particulares, para poder dar respuestas a las necesidades que les plantean.

Por otra parte la viabilidad del proyecto tampoco estuvo totalmente clara al principio. La mayor parte de la documentación estaba en inglés y se encontró poca información en castellano. En cualquier caso está duda se desvaneció al encontrar la herramienta treeTagger y ver que posibilitaba el análisis de textos en cualquier idioma, facilitando el trabajo de extraer y clasificar las diferentes palabras que forman el tweet.

Siguiendo con la línea de treeTagger, se tendría que revisar el corpus puesto que en algunas ocasiones, al analizar las palabras, desde mi punto de vista, no cuadran completamente con el tipo de palabra al que se asocia.

Otro punto a destacar, sería el API de twitter, en este caso se ha observado que, al hacer la consulta para conseguir los tweets, en algunas ocasiones salen duplicados o triplicados. Esto ocurre por el número de retweets que se han realizado a ese tweet, es una cosa que no se puede controlar puesto que al realizar la llamada al API devuelve los tweets más recientes sin analizarlos. Por lo que se puede plantear quitar los retweets y solo mostrar uno, esto puede producir una pérdida de información pero se piensa que puede ser mejor solución.

Para concluir hay que mencionar, que últimamente se han encontrado diferentes empresas españolas que se publicitan para que las personas que lo deseen, puedan saber el análisis de datos para empresas.



Análisis de sentimientos acerca de marcas mediante Tweets

Por tanto como conclusión global se puede decir que este proyecto podría ayudar a mucha gente con su trabajo; no solo con el tema que nos atañe de *community manager*, si no, qué también puede ser utilizado para mejorar o ayudar a mejorar la calidad de vida a personas discapacitadas.



CAPÍTULO 8

Trabajo Futuro

Este capítulo se puede explicar que mejoras se le pueden realizar. Estas mejoras se les han ocurrido a diseñadores y programadores, los cuales creen que podrían favorecer la utilización de la aplicación.





8 TRABAJO FUTURO

En esta sección se van a mostrar las futuras líneas de investigación, las cuales han aparecido a lo largo del proyecto y que se creen que podrían ayudar en gran medida al usuario que utilice la aplicación, en este caso al *community manager*.

La primera opción sería el poder **Guardar las Búsquedas Realizadas**. Esta opción se puede realizar de tres formas.

1. Guardar todas las búsquedas en la aplicación. Si se opta por esta opción la página inicial sufriría modificaciones. Se generaría una nueva pestaña “Historial” en la que se guardarían las búsquedas por marca, cada una en una subpestaña, y estas se ordenarían por fecha. Se generarían pdf o excel.
2. Poder descargar las búsquedas directamente al ordenador. Simplemente se añadiría un botón debajo de la explicación grafica total y si al manager le interesa la descargaría sino no se quedaría guardada.
3. Unificarlas. Esta opción sería la más óptima, puesto que si por un casual el manager no realiza una copia física en su ordenador, y más adelante la necesita siempre podría volver a la aplicación para poder recuperarlo, las búsquedas se guardarán automáticamente en la aplicación.

La segunda opción sería el **Crear Nuevos Filtros**.

En el momento en el que se cambie el idioma de la página, aparte de cambiar en apariencia, lo que se realizaría sería buscar los *tweets* en el idioma elegido.

Otros filtros que se podrían incluir serian: Buscar *tweets* de personas comprendidas entre dos edades, Buscar *tweets* dentro de un periodo de tiempo específico (esto se podría realizar dando cuadros de texto para escribir las fechas, crear unos por defecto: un día, una semana, quince días, un mes) o dar las dos opciones e inhabilitar la contraria por la que se quiere buscar.

La tercera opción sería **Introducir gráficos**. Estos gráficos se van a mostrar de diferentes formas.

- 1- El primer gráfico se podría utilizar para mostrar el número de *tweets* que se han obtenido en cada búsqueda y estos agruparán por cada día de la semana.



Este grafico se verá modificado según el número de búsquedas que se realicen al día, para que tenga lógica se tendrán que realizar el mismo número de búsquedas cada día.

Una vez que se tengan todos los resultados se dispondrán en un diagrama que mostrará con qué frecuencia se ha nombrado a la marca en los *tweets*.

- 2- El segundo gráfico ya sería más completo, en este caso se mostrarían los *tweets* divididos por los tres grupos que se han mostrado siempre: favorables, neutros y desfavorables, pero en este caso si presionamos sobre alguno de esos grupos, aparecerá otro grafico en este caso definido por edades.

Con esta nueva división se da un poco más de información al *community manager* para que sepa en qué tipo de personas hay que hacer mayor hincapié a la hora de poder realizar campañas de publicidad para cambiar ese malestar.



CAPÍTULO 9

Referencias

Este capítulo se va a disponer todas las páginas o sitios webs en los que se han entrado para poder obtener la información que se ha ido disponiendo en este documento.

Por otro lado también se pueden encontrar diferentes artículos que han sido leídos para poder comprender el sentimiento de análisis de los cuales se ha podido extraer solo conocimiento para poder explicar otros términos o información que ha sido plasmada.





9 Bibliografía

- Bing, Liu . (s.f.). *Sentiment Analysis and Subjectivity*, 4.
- Bing. (s.f.). *Función near*. Obtenido de <http://cosascuriosasdelainformatica.blogspot.nl/2011/08/el-operador-near-de-bing.html>
- Bosque, I., & Guitiérrez-Rexach, J. (2009). Fundamentos de Gramática Formal .
- Buenavista, I. (03 de marzo de 2009). Facebook se queda con los derechos de autor de nuestras fotos. *Antena 3 noticias*, p. D01.
- Cortizo, J. C. (13 de Mayo de 2011). *Minería de Opiniones, o Análisis del Sentimiento*. Obtenido de <http://www.baquia.com/posts/2011-05-13-mineria-de-opiniones-o-analisis-del-sentimiento>
- Encyclopedia, B. O. (2002). Markov process (mathematics). En B. O. Encyclopedia, *Markov process* .
- Estocásticos, M. (2006). *Modelos estocásticos*. Obtenido de http://www.itch.edu.mx/academic/industrial/sabaticorita/_private/07Procesos%20estocasticos.htm
- Fernández, J., Boldrini, E., Gómez, J. M., & Martínez-Barco, P. (2010). *El corpus EmotiBlog*. Obtenido de <http://journal.sepln.org/index.php/pln/article/view/963>
- Freeman, L. (2006). The Development of Social Network Analysis.
- Gómez Hidalgo, J. M. (s.f.). *Procesamiento lenguaje natural*. Obtenido de <http://www.esi.uem.es/~jmgomez/pln/index.html>
- Google. (s.f.). *función around*. Obtenido de <http://googlesystem.blogspot.nl/2010/12/googles-around-operator.html>
- Halliday, M. (1975). Estructura y función del lenguaje. Alianza .
- HOMSANI, S. (2000). Etiquetados de textos con Conexor.
- Iliffe, P. (14 de noviembre de 2011). «Tuenti: la red social más privada, cada vez más móvil.
- Melber, A. (12 de febrero 2008). Does Facebook Own You Forever? *The Huffington Post*.



- Millán, S. (12 de mayo de 2011). Tuenti eleva su plantilla más de un 36% desde enero. *Cinco Días*.
- Mitchell. (1997). T.M. Machine Learning. McGraw-Hill.
- Monsalve, B. (13 de diciembre de 2010). *Analisis del sentimiento y mineria de opiniones*. Obtenido de <http://www.brainsins.com/es/blog/analisis-del-sentimiento-y-mineria-de-opiniones/99679>
- Murphy, K. (s.f.). Hidden Markov Model (HMM) . *Toolbox para Matlab* .
- Natural, A. M. (30 de octubre de 2009). *Procesamiento del Lenguaje Natural*. Obtenido de <http://www.cicling.org/ampln/index-tmp.htm>
- RH, R. (1989). Part of Speech. Londres: General Linguistics.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*.
- Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*.
- Sentisis. (2013). *Sentisis*. Obtenido de <http://www.sentisis.com/about-us-technology>
- Simón, A. (s.f.). Tuenti, la competencia de Facebook hecha en España. *Cinco Días*.
- Twitter. (2010). *¿Qué es un tweet?* Obtenido de <http://myspace.wihe.net/que-es-un-tweet/>
- Twitter. (s.f.). *Api de Twitter*. Obtenido de <https://dev.twitter.com>
- Ureña, L. A., & Martínez, F. (2011). Tratamiento de la Información Multilingüe y.



Anexo I

Glosario de Términos

Este anexo se encuentra ciertas palabras o conceptos que pueden ser ajenos al lector. Por tanto la función que desempeña esta parte es la de explicar lo mejor posible cada uno de los términos para que al lector le sea más cómoda la lectura y por tanto mayor comprensión sobre el tema que se trata en este documento.





Anexo I. GLOSARIO DE TÉRMINOS

Red Social: SON estructuras sociales compuestas de grupos de personas, las cuales están conectadas por uno o varios tipos de relaciones, tales como amistad, parentesco, intereses comunes o que comparten conocimientos

App: aplicación

Tuenti: red social española, creada en 2006 que cuenta con más de 14 millones de usuarios. Permite al usuario crear su propio perfil, añadir a otros usuarios como amigos e intercambiar mensajes, fotos y vídeos, páginas o eventos. Tiene servicio de chat (individual y en grupo) y videochat (solo con una persona).

Facebook: Originalmente era un sitio para estudiantes de la Universidad de Harvard, pero actualmente está abierto a cualquier persona que tenga una cuenta de correo electrónico. Los usuarios pueden participar en una o más redes sociales, en relación con su situación académica, su lugar de trabajo o región geográfica.

Twitter: La red permite enviar mensajes de texto plano de corta longitud, con un máximo de 140 caracteres, llamados *tweets*, que se muestran en la página principal del usuario.

Tweet: Tweet es algo así como un mensaje entre pájaros, aunque en la red social significa más bien una publicación o actualización en el estado de un usuario.

Communitiy Manager: Es la persona responsable de la comunicación de una empresa, la cual se encarga de crear, gestionar y dinamizar una comunidad de usuarios en internet.

Postear: Es la acción de publicar un *tweet* en tu muro o en el de otra persona.

Post: Es como se denomina la *tweet* ya publicado en tu muro o en el de otra persona.

Microblogging: Es la tecnología en la que se basa *twitter* para poder funcionar, es expresar lo que se piensa en un numero definido de caracteres., sin poder sobrepasarlos.

Developer: Palabra que se utiliza para denominar a un desarrollador software.



Interface: Pantalla que se le mostrara al usuario, en la que éste puede interactuar con la aplicación.

Feelings: Sentimientos



Anexo II

Manual del Usuario

Este anexo contiene un pequeño manual para que el usuario sepa cómo utilizar la aplicación.





Anexo II. MANUAL DEL USUARIO

En este anexo se explicará de qué forma tiene que utilizar el usuario la aplicación para que pueda realizar el análisis de la marca que desee.

Para que todo funcione de forma correcta hay que descargar el *TreeTagger* de la siguiente página web.

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

En la página web, anteriormente mencionada, da todas las indicaciones correspondientes para poder instalar y ejecutar todo lo relacionado con el analizador de textos *TreeTagger*. Al estar en inglés, se traducirán los primeros pasos a seguir para poder descargarlo.

- 1- Descargar el paquete de *treetager* para su sistema operativo. ([PC-Linux](#), [Mac OS-X \(Intel-CPU\)](#), [PC-Linux \(version for older kernels\)](#)).
- 2- Descargar los [tagging scripts](#) del mismo directorio.
- 3- Descargar los script de instalación [install-tagger.sh](#).
- 4- Descargar los ficheros de parámetros según el sistema. ([PC](#), [Mac-Intel](#)).
- 5- Abrir un terminal para poder ejecutar el script de instalación en el directorio en el que se han descargado los ficheros. Y ejecutar:
`sh install-tagger.sh`

Make a test, e.g. `echo 'Hello world!' | cmd/tree-tagger-english` or
`echo 'Das ist ein Test.' | cmd/tagger-chunker-german`

El primer paso, será abrir la aplicación *Análisis de sentimientos acerca de marcas mediante Tweets*. La 0 muestra cómo tiene que aparecer la aplicación al ser ejecutada por primera vez. Antes de comenzar a explicar el segundo y tercer paso, se indicará que valores pueden tener las diferentes etiquetas que muestra la aplicación.

- Las tres etiquetas “Resultado” cambiarán por el número de *tweets* de ese sentimiento en concreto.
- La etiqueta “En espera” tendrá tres valores diferentes, en función de en el momento en el que se encuentre el proceso. Los valores pueden ser “En espera” antes de empezar la ejecución, “En ejecución” en el momento en el que empiezan la ejecución, “Finalizado” en el momento en el que se acaba la ejecución.
- La etiqueta “Resultado final” podrá dar tres tipos de frases diferente:
 - o Resultado negativo, el separador estará en el primer tercio de la barra junto con la frase: “Malos sentimientos hacia la marca”.



Análisis de sentimientos acerca de marcas mediante Tweets

- Resultado neutro, el separador estará en el segundo tercio de la barra junto con la frase: “La marca no es atractiva”.
- Resultado positivo, el separador estará en el tercer tercio de la barra junto con la frase: “Buenos sentimientos hacia la marca”.

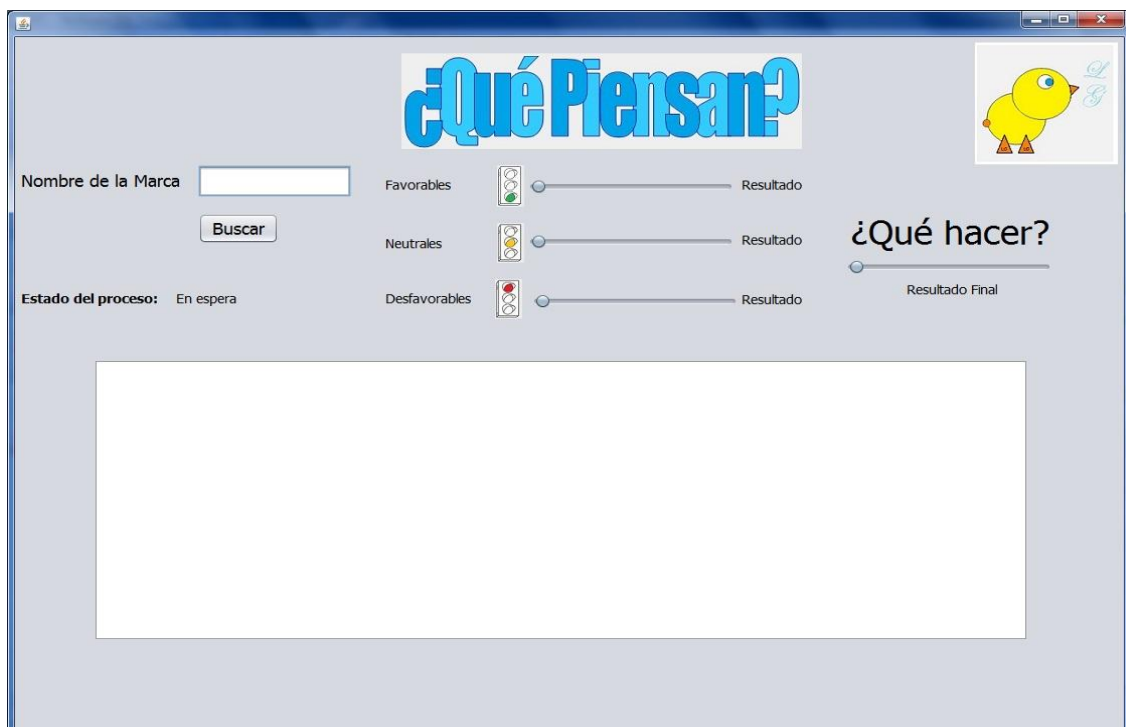


Ilustración 46. Primera pantalla

En la **Ilustración 47** se muestran los pasos que hay que seguir para poder utilizar la aplicación.

- 1- Introducir el nombre de la marca, la cual se quiere analizar.
- 2- Pulsar el botón buscar.
- 3- El estado del proceso pasará de “En espera” a “En ejecución”

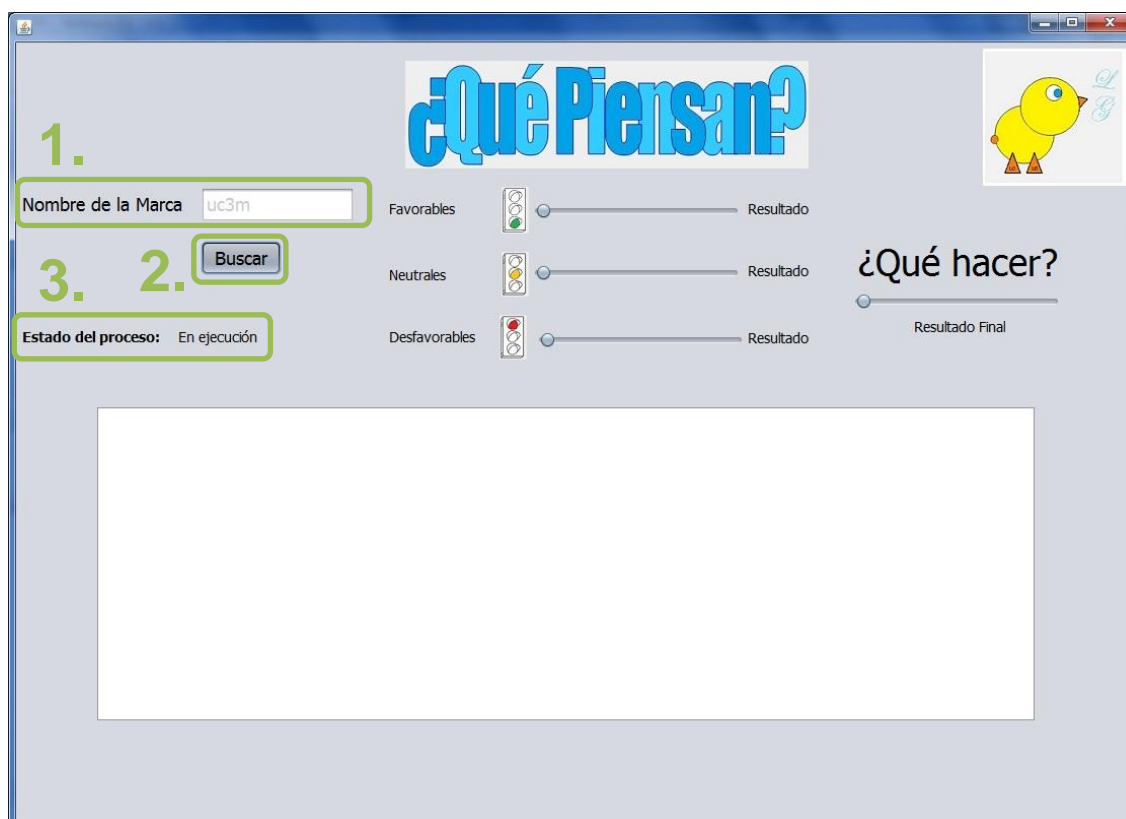


Ilustración 47. Segunda pantalla

En la **Ilustración 48** se muestran los pasos que hay que seguir para poder utilizar la aplicación.

- 1- El botón *Buscar* cambia de color.
- 2- El estado del proceso pasará de “En ejecución” a “Finalizado”.
- 3- Se muestran todos los *tweets* que se han encontrado con sus sentimientos a continuación de cada uno de ellos. Al final de todos ellos aparecerá el sentimiento global de la marca.
- 4- Se desplazarán las barras hasta el número que indiquen los números que las acompañan.
- 5- La barra indicará el sentimiento global de la marca junto con una frase explicativa.

Nombre de la Marca

2

Estado del proceso: Finalizado

¿Qué Piensar?

Favorables 3

Neutrales 20

Desfavorables 2

¿Qué hacer?

La marca no es atractiva

3

Con el tweet: "Eres más borde que los trabajadores del copyred de la uc3m" #concepto #uc3m su sentimiento es: Neutro
Con el tweet: "Saciado Morfeo con 1 hora más no hay excusa para perderse #mpmTwitterTour y cantar el http://t.co/5m4L7VswUi a @mPICASSOm Entona @mediamus
Con el tweet: RT @carolinare: #esnmadness y @esn_uc3m me han dado más de lo que nunca podre devolverles. Gracias por convertirme en quien soy, su sentim
Con el tweet: RT @MGC_UC3M #Convocatoria de Becas 2014 del Museo Nacional Reina Sofia. Madrid http://t.co/9H2Ww0Jeki #GestiónCultural su sentimiento es: N
Con el tweet: @agaron Precios del Master de Abogacía (obligatorio para ejercer): 1200 € en UMA, 6750€+gastos administrativos en UC3M. Educación "pública" su se
Con el tweet: Ojo a la #Convocatoria de Becas 2014 del Museo Nacional Reina Sofia. Madrid http://t.co/3Vyz18jDZp #GestiónCultural su sentimiento es: Neutro
Con el tweet: @charlies_uc3m Quieres formar parte del equipo de RRPP Discoloca? Geniales condiciones! +Info por Whatsapp: 658121009 :) su sentimiento es: Ne
Con el tweet: RT @jcarolinare: #esnmadness y @esn_uc3m me han dado más de lo que nunca podre devolverles. Gracias por convertirme en quien soy, su sentim
Con el tweet: Viva la uc3m y su envío de notas un sábado en la noche; su sentimiento es: Neutro
Con el tweet: #esnmadness y @esn_uc3m me han dado más de lo que nunca podre devolverles. Gracias por convertirme en quien soy, su sentimiento es: Neutro
Con el tweet: RT @Mario_Colleoni: @MGC_UC3M @espaciolabruc Me mandaron la invitación pero finalmente no pude acercarme. ¡Tengo muchas ganas de conocer
Con el tweet: @Mario_Colleoni @MGC_UC3M Las puertas abiertas para cuando queráis! su sentimiento es: Bueno
Con el tweet: 3 días más, solo 3 días más!!! #gogo #uc3m #PFM #MSc su sentimiento es: Neutro
Con el tweet: @Nuria_Onoro ¿Se refiere a la petición de 2 millones de pesetas por parte de los Mateo? El TEDH lo rechazó http://t.co/MyCU4cJWbb su sentimiento es:
Con el tweet: RT @QueenCarlosIII: Si os sentís tristes al ir a la uc3m, es porque han contratado dementores. su sentimiento es: Malo
Con el tweet: RT @sobreviveuc3m: Vídeo motivador de cara a los exámenes de Junio 2013. http://t.co/VQIOBTOKg #uc3m su sentimiento es: Neutro
Con el tweet: RT @ferre305: Q nos entrw en las margaritas to la #UC3M de fiesta borrachos n el tren no tiene precio @MarcosHTwitt su sentimiento es: Neutro
Con el tweet: Q nos entrw en las margaritas to la #UC3M de fiesta borrachos n el tren no tiene precio @MarcosHTwitt su sentimiento es: Neutro
Con el tweet: RECORDATORIO beUp! Esta noche se cambia la hora!! @UAHes @ucomplutense @uc3m @UAM_Madrid @UBUEstudiantes @UDC_gal http://t.co/Mr

Ilustración 48. Tercera pantalla